SYMMETRY, DYNAMICS AND FUNCTION: BIOLOGICAL MACROMOLECULES STUDIED BY ELASTIC NETWORK MODELS

by

Zheng Yang

B.S. in Physics, Fudan University, China, 2003

M.S. in Physics, University of Pittsburgh, 2007

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

Department of Physics & Astronomy, School of Arts & Science

This thesis was presented

by

Zheng Yang

It was defended on

July 29, 2009

and approved by

Ivet Bahar, Department of Computational Biology, School of Medicine

Rob Coalson, Department of Physics & Astronomy, School of Arts & Science

David Jasnow, Department of Physics & Astronomy, School of Arts & Science

Joseph Boudreau, Department of Physics & Astronomy, School of Arts & Science

Vladimir Savinov, Department of Physics & Astronomy, School of Arts & Science

Thesis Advisor: Ivet Bahar

Thesis Co-Advisor: Rob Coalson

SYMMETRY, DYNAMICS AND FUNCTION:

BIOLOGICAL MACROMOLECULES STUDIED BY ELASTIC NETWORK MODELS

Zheng Yang, Ph.D.

University of Pittsburgh, 2009

Symmetry is a common feature in nature. Large biological macromolecules (> 100 kD) tend to assemble from multiple subunits and spatially arranged in symmetric ways. The topological symmetry not only results in coding efficiency and error control, but also characterizes the equilibrium dynamics of the biomolecular system. Coarse-grained normal mode analyses have been broadly used in recent years to elucidate the relation between structure, dynamics and function. Further insights into collective motions can be gained by considering continuum models with appropriate symmetry and boundary conditions to approximately represent the molecular structure. We solved the elastic wave equation analytically for the case of spherical symmetry, yielding a symmetry-based classification of vibrational motions accessible to the structures together with explicit predictions of their vibrational frequencies. Applications to biomolecular assemblies have shown that the continuum models with spherical symmetry efficiently provide insights into collective motions that are otherwise obtained by detailed elastic network models. Additionally, to understand the mechanism of functions associated with structural changes between different conformations, the transition pathways between these conformations have been explored with the help of elastic network models. Although there are many computational methods for exploring the conformational transitions of proteins, these are usually applicable to small-to-moderate size proteins, and the task of exploring the transition

pathways becomes prohibitively expensive in the case of supramolecular systems of the order of megadaltons. Coarse-grained models that lend themselves to analytical solutions appear to be the only possible means of approaching such cases. Motivated by the utility of elastic network models for describing the collective dynamics of biomolecular systems, and by the growing theoretical and experimental evidence in support of the intrinsic accessibility of functional substates under native state conditions, we developed a new method, *adaptive anisotropic network model* (*a*ANM), for exploring the functional transitions of large biomolecular systems. Application to bacterial chaperonin GroEL, and comparisons with experimental data, and with results from other theoretical and/or computational approaches, support the utility of *a*ANM as a computationally efficient, yet physically plausible, tool for unraveling the potential transition pathways sampled by large complexes/assemblies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Biological macromolecules are at the center of action in biological processes, such as gene expression, signal transduction and viral replication [1]. In this chapter, we will present an overview of macromolecular structural symmetries, dynamics modeled by elastic network models, functional movements and the related states transition problem.

1.1 SYMMETRY AND BIOLOGICAL MACROMOLECULES

Starting from the first X-ray structure of a protein, sperm whale myoglobin, by John Kendrew and co-workers (1958) [2], there have been 57,706 structures deposited into the Protein Data Bank (PDB) [3] (as of 05/21/2009). A considerable fraction of the structures reveal different symmetry properties at various levels. For example homodimers (proteins composed of two identical subunits, or monomers) are very common, and usually the two subunits are symmetrically arranged with respect to a symmetry axis.

Based on their one-dimensional sequence of amino acids, polypeptides usually fold into regular patterns, such as helices and sheets, which are elements of the tertiary structures. Most proteins, especially large proteins (>100 kD), are assemblies of multiple chains, also called multimers, or oligomers. The preference for oligomeric proteins may be due to error control [4], coding efficiency and regulation of assembly [5, 6]. Additionally, the homooligomeric proteins are highly symmetrical, their multiple subunits being arranged by simple point groups. Due to

the chirality of the amino acids, inversion or mirror symmetry are prohibitted. But rotational symmetries are favorable such as cyclic, dihedral, tetrahedral, octahedral and icosahedral symmetry.

The symmetric arrangement of subunits is essential for structural stability [7], subunits assembly, and folding efficiency [8]. More recent studies also point to the importance of symmetry for achieving cooperative responses [9], in accord with the Monod-Wyman-Changeux (MWC) model for allostery [10, 11]. Accordingly, all subunits simultaneously undergo the same type of structural change in response to ligand binding or other external perturbation, and the symmetry of the structure facilitates the cooperative response of the intact structure. The alternative model of allostery, known as the Koshland-Némethy-Filmer (KNF) model assumes that the changes in the individual subunits occur sequentially [11, 12]. Whether the MWC or KNF mechanism controls allosteric responses is still an open question, and many allosteric systems may exhibit a mixed behavior.

Our recent study of mechanism of allostery in bacterial chaperonin GroEL, an ATP-regulated molecular machine composed of two heptameric rings, suggests that the dynamics intrinsically favored by the 7-fold symmetric architecture is coupled to the protein's allosteric signal transduction pathways [13]. In the following two chapters, the relation between structural symmetry and dynamics will be explored. We will focus on icosahedral symmetry (denoted by I) in the second chapter. Viral capsids with icosahedral symmetry will be approximated as continuum models of spherical symmetry to examine their elastic properties. In the third chapter, the transition pathways between multiple intermediates will be explored. A novel pathway generating method will be applied to the allosteric transitions of GroEL, which has a 7-fold rotational/dihedral symmetry (denoted as D_7). It will be shown that the most productive

pathways enabling the passage between different functional forms of the chaperonin are those preserving the rotational symmetry of the overall structure and incurring minimal redistribution in native contacts.

1.2 DYNAMICS AND ELASTIC NETWORK MODELS

1.2.1 Elastic network models

Recent years have seen a revival in normal mode analyses (NMAs) of proteins, after realizing that coarse-grained models such as the elastic network models (ENMs) provide information on functional and robust modes [14-18]. ENMs allow for efficient calculation of the global motions of biomolecules, making them a suitable complement to conventional molecular dynamics (MD) simulations, especially when exploring the equilibrium dynamics of large molecular systems. The elastic network models used in coarse-grained NMA include: the one-parameter Gaussian Network Model (GNM) [19] and Anisotropic Network Model (ANM) [20, 21]; the rotational translational block (RTB) [22-24] or block NMA (BNM) [25]; hierarchical ANM [26], hierarchical clustering based on Markovian stochastics [27], and several other variations such as mixed elastic network models [28], β-ANM where the network nodes are located at the β-carbons, rather than α-carbons [29]. Such reduced models proved useful for exploring the complex machinery of supramolecular systems such as the ribosomal complex [30, 31], GroEL-GroES [13, 32, 33], viral capsids [34-36], or large structures determined by cryo-electron microscopy (cryo-EM) [37-40].

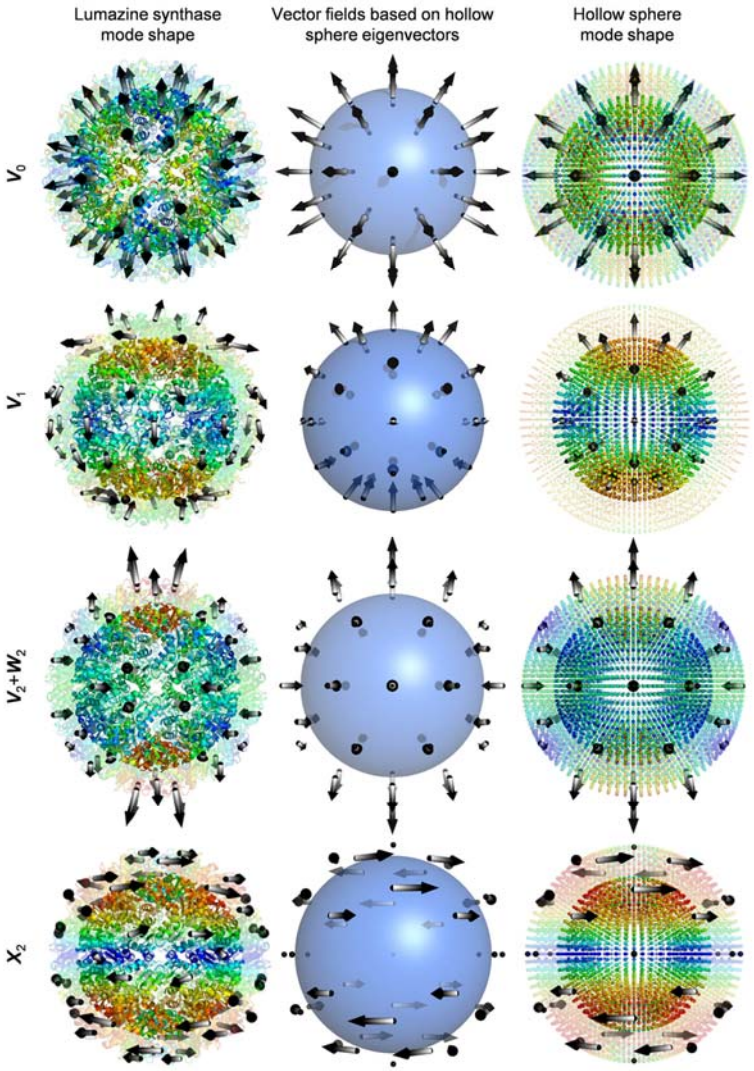

Figure 1. Correspondence between ANM modes and vector spherical harmonics illustrated for lumazine synthase (LS).

The diagrams on the left display the structure (from PDB file 1NQW [41]) color-coded according to the mobilities of residues predicted by the ANM, from blue (small) to red (large). The arrows indicate the directions of collective motions on or above the *middle surface*, halfway between the inner and outer radii. The interior (inside the middle surface) portion is shown in opaque, the exterior portion in transparent. The middle and right column display the equivalent vector fields and mode shapes generated by the ANM for a hollow sphere with comparable dimensions and packing density. The diagrams on the right column are also colored by the mobilities of the nodes. The labels on the left indicate the vector spherical harmonics corresponding to the displayed modes.

These studies suggest that molecular shape governs dynamics, especially in the low frequency regime [42]. Indeed one can gain insights into the collective motions relevant to biological function by examining the lowest frequency modes predicted by NMA conducted either at the atomic level, or using low resolution models [14, 43, 44]. A subset of low frequency modes sampled by symmetric structures are themselves symmetric in character, thus maintaining the symmetry of the original structure (see Figure 1).

1.2.2 Anisotropic network model

In this thesis, the ANM will be utilized to examine various biological systems. Based on the ANM, an adaptive method will be built to sample entropically favored transition pathways between intermediate states.

The basic approach in the ANM analysis of proteins is to determine the ensemble of normal modes of motion accessible to the examined structure under equilibrium conditions, assuming that the close neighborhood of the energy minimum can be approximated by a harmonic potential. Conventional NMA requires energy minimization prior to eigenvalue decomposition of the $3N \times 3N$ Hessian matrix representative of the second derivatives of the potential with respect to the components of the $3N$ generalized coordinates that define the structure. In the ANM, the nodes of the elastic network are typically defined by the C_α atom coordinates, and the springs are representative of the inter-residue interactions within a cutoff distance r_c [20, 21]. The native structure deposited in the PDB is assumed to be the equilibrium state such that no further energy minimization is required.

In particular, the ij^{th} super-element (3×3 matrix) of the Hessian takes the form [21]

$$\mathbf{H}_{ij} = \frac{\gamma\Gamma_{ij}}{(R_{ij}^0)^2}\begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix}, \tag{1.1}$$

which directly follows from the adoption of a harmonic potential with uniform force constant γ,

$$E = \frac{\gamma}{2}\sum_{j|j\neq i}\Gamma_{ij}(R_{ij} - R_{ij}^0)^2. \tag{1.2}$$

Here R_{ij} and R_{ij}^0 are the respective instantaneous and equilibrium distances between nodes i and j. $\mathbf{\Gamma}_{ij}$ is the ij^{th} element of the Kirchhoff matrix $\mathbf{\Gamma}$ of inter-residue contacts, set equal to 1 if R_{ij}^0 is smaller than a cutoff distance r_c, zero otherwise [19, 21]. X_{ij}, Y_{ij}, and Z_{ij} are the components of the distance vector R_{ij}^0. The diagonal elements of $\mathbf{H}$ are defined as the negative sum of all off-diagonal terms in a given row (or column; $\mathbf{H}$ is symmetrical).

The squared frequency and shape of each mode is given by the nonzero eigenvalue (λ_k) and corresponding eigenvector ($\mathbf{v}^{(k)}$), respectively, of $\mathbf{H}$. We denote the nonzero eigenvalues of $\mathbf{H}$ as λ_k $(1\leq k\leq 3N-6)$ in ascending order, and the corresponding eigenvector as $\mathbf{v}^{(k)} = \left[\mathbf{v}_1^{(k)},\ldots,\mathbf{v}_N^{(k)}\right]^T$, where $\mathbf{v}_i^{(k)}(1\leq i\leq N)$ is the 3-dimensional displacement vector of residue i induced by mode k. We are interested here in the lowest frequency portion of the mode spectrum, which we obtain by Arnoldi iteration [45, 46].

1.3 FUNCTION AND MOTIONS

1.3.1 Proteins in motion

It is now well established that proteins are in constant motion *in vivo*, sampling an ensemble of configurations [47-50]. Additionally, by undergoing changes between different conformations, proteins carry their biological functions. Understanding the mechanism of transition between structures is of major importance to designing methods for controlling such transitions, and thereby modulating protein functions.

However, exploring the transition between conformations is hard, both experimentally and computationally, due to the transient nature of the intermediate, high energy conformers crossed over as the molecule undergoes structural changes. In many cases, only the two ending structures are known from experiments [3]. Furthermore, the passage between the two end points does not necessarily involve a single pathway, but multiple pathways in the multidimensional energy landscape associated with the macromolecular structures. To bridge between structure and function, a molecular understanding of the most probable transition pathways between the two end structures is required [51, 52].

1.3.2 Computational methods for exploring transition pathways

A broad range of computational methods have been developed in the last two decades for exploring the transition pathways between the different functional forms of biomolecules. In their pioneering studies, Elber and Karplus (1987), and Czerminki and Elber (1990) proposed to generate plausible reaction pathways upon minimization of path-dependent functionals [53, 54].

Olender and Elber (1996) used a stochastic path integral to integrate classical Newtonian dynamical equation of motion [55]. Huo and Straub developed the MaxFlux method, where the optimal reaction path was defined as the path of maximum flux (1997) [56], a topic recently revisited [57]. A modified method to find optimal path, named nudged elastic band method, was proposed by Jónsson, Mills and Jacobsen (1998) [58]. Methods for global optimization of the functionals that define the paths have been developed that take into account the temperature of the system. Elber and Shalloway determined temperature-independent reaction coordinates by minimizing the action in the lower limit of temperature [59]. The idea of simulating reaction coordinates via action optimization, as opposed to a classical solution of Newton's equation has been promoted by Elber's group [60]. For more details, see a recent review on long-timescale simulation methods by Elber [61]. The detailed derivation of the steepest descent path (SDP) [62] is given in Appendix B.

In addition to these physically based analytical approaches, a number of researchers explored the transition pathways by simple interpolation between known conformers. A study along these lines is that of Vonrhein *et al* (1995) who interpolated between 17 solved nucleoside monophosphate kinase structures to explore the structural changes occurring during their catalytic cycle [63]. More extensive usage of interpolation is the database of molecular movements MolMovDB developed by Gerstein lab; MolMovDB is based on morphing between known end points [64].

In another group of studies, targeted molecular dynamics (TMD), proposed by Wollmer and coworkers (1993), has been performed for studying the transition problem [65]. In TMD, the system is simulated in the presence of a force field and holonomic constraints. Ma and coworkers applied the TMD method to a series proteins, such as *ras* p21 (1997) [66], GroEL (2000) [33],

F1-ATPase (2002) [67], and lactose repressor protein (2003) [68]. Zhang *et al* utilized the TMD on human glucokinase (2006) [69]. Similarly, steered molecular dynamics (SMD) could reveal the process under atomic force microscope (AFM) pulling/pushing, to study the process of unfolding and interaction with ligand or ions (1996) [70]. Recently, a beautiful study by Gu *et al*, applying the SMD on glutamate transporters, elucidated the substrate translocation pathway (2009) [71].

To generate sufficient transition events, especially rare events, rather than wasting computing resources on the waiting time between events, Monte Carlo (MC) path sampling approach has been adopted. It was initially Pratt who applied the method to locating the transition states of chemical systems (1986) [72]. Chandler and coworkers improved the algorithm efficiency to allow for the calculation of rate constants in complex molecular systems (1998) [73, 74]. Zimmer and Paniconi characterized the statistical properties that describe the growth of large fluctuations for a bistable stochastic system (1999) [75].

Also based on the idea of importance sampling like MC, Ottinger introduced a compensating bias in favor of those configurations which mainly contribute to the average of a given quantity of interest(1994) [76]. This approach considerably reduces the variance of the stochastic simulation results. Zuckerman and Woolf followed the same route and developed the dynamic importance sampling, which was demonstrated for a single particle moving on the two-dimensional Müller–Brown potential surface (1999) [77-79].

Huber and Kim developed a weighted-ensemble method, by assigning different weights to multiple simulations that were simultaneously conducted and by distributing computing resources, to efficiently sample the rare events (1996) [80]. Rojnuckarin and coworkers applied the method on the simulations of a four-helix bundle protein folding (1998) [81], bimolecular

association problems (2000) [82], kinetics of dimeric hemoglobin (2001) [83], and the effects of point charge mutations on the electrostatically driven association of hemoglobin subunits (2002) [84]. Zhang *et al* utilized the weighted ensemble method to study a large-scale transition in a united-residue model of calmodulin (2007) [85]. The results were validated by comparing them with brute-force simulations. In the method, a previous chosen reaction coordinate was divided into multiple bins. Each bin was populated and contained several independent simulations. M independent trajectories were initiated from the same configuration with equal weight, 1/M. After running the simulation for a period time, each trajectory was divided or combined with others based on the criterion that the total number of trajectories inside each bin must be either M or 0; and the trajectories weight was added or divided according to splitting or combining operation respectively, to ensure the normalization of the weights. The process was repeated until a representative ensemble of trajectories was obtained.

Yet, the identification of the transition states and accompanying conformational rearrangements are largely inaccessible via computations for systems of the order of megadaltons like GroEL. Coarse-grained models and methods appear as the only tractable tools in such cases.

1.3.3 Elastic network models for exploring transition pathways

It has been shown that the transitions between the initial and final states of a given molecular system are not likely to be along a single linear trajectory. Instead, there is an ensemble of paths or trajectories moving on the surface of a complicated multidimensional energy landscape. Recent studies indicate that the global transitions proceed, or at least start, via the collective global mode directions that are predicted by NMA, and it has been shown in many applications that conformational changes may be accounted for to a good approximation in terms of a few

slow modes [18, 43, 86]. This observation puts ENMs amongst the tools for theoretical studies of transition pathways. Since NMA is valid only in the local region surrounding a potential energy minimum, its application to non-equilibrium events such as conformational changes must be handled delicately.

Kim *et al.* initially proposed to interpolate between the end points by controlling the *distances* between the nodes with ENM, rather than the position vectors so as to avoid unrealistic deformation of the structural geometry (2002) [87, 88]. In this method, the system was first modeled as an EN for the initial state, and the transition proceeded by gradually reducing the effects of the initial state's EN while increasing the contribution of the final state's EN. Throughout the transition, intermediate conformations were calculated by minimizing a cost function based on the instantaneous EN. Three years later, this method was further developed to build the rigid-cluster ENM (2005) [89].

Zheng and Brooks constructed an ENM for the initial state only, and then slowly perturbed the structure to satisfy known distance constraints from the final structure (2005) [90]. The method has been tested on 16 pairs of protein structures, which have sufficiently distinctive structures (2006) [91]. Onuchic and coworkers extended the concept of minimal frustration to transitions between basins modeled as ENMs, and examined the coupling between the strains in dihedral angles and local unfolding events (also termed *cracking*) [92, 93]. Putative transition pathways could be generated in these methods with a relatively low computational cost.

Another technique, so-called plastic network model, was developed by Maragakis and Karplus, and applied to the transition of adenylate kinase between its T and R forms. The transition states were generated therein by using a combination of steepest descent and conjugate peak refinement algorithms (2005) [94]. A similar 'mixed ENM' was used to investigate the

helix-to-sheet transition of Arc repressor (2005) [95]. Chu and Voth modified the plastic network model and presented a double-well ENM model for exploring the conformational transitions of G-actin and adenylate kinase (2007) [96].

Alternatively, by taking advantage of the ability of ENMs to provide an understanding of the global movements undergone by large systems, and toward compensating for the loss of accuracy and specificity at the local scale, Isin *et al* developed a hybrid approach, ANM-steered MD. The idea is to steer MD trajectories along ANM modes. The method has been successfully applied to studying the photoactivation dynamics of rhodopsin (2008) [97]. In this method, global conformational changes that are not accessible via conventional MD trajectories can be sampled, while motions and interactions at atomic scale can be observed in the presence of explicit solvent and lipid bilayer.

Franklin *et al* built a web server, MinActionPath, to generate transition trajectories that are sampled along the maximum likelihood path (2007) [98]. In their method, the classical C^{α} based ENM is adopted for both the initial and final states. By using Onsager and Machlup action minimization formalism [99], the Langevin equations were analytically solved. Although limited to 1,000 atoms, this method yields a relatively fast analytical solution, by reducing the non-linear problem to a pair of linear differential equations joined by a non-linear boundary matching condition.

It is worth noting that not all NMA-based methods of generating transition pathways are strictly analytical. The MC Normal Mode Following (MC-NMF) method, developed by Miloshevsky and Jordan, utilizes normal modes to guide the simulations (2006) [100]. In their method, the initial structure was perturbed moving energetically uphill along the lowest frequency normal mode, while simultaneously minimizing the energy along all other modes. The

size of the deformation was arbitrary; and the acceptance of a given move was based on the Metropolis criterion [101]. After each accepted move, the Hessian matrix was recalculated and rediagonalized. Multiple runs yield an ensemble of transition states structures. Gramicidin A and K^+ channel from *Streptomyces lividans* (KcsA) [102] were examined with the MC-NMF method. An advantage of this technique is that it does not require two conformations, but generates an approximate potential energy landscape from the initial conformation alone.

The success of these studies suggests that ENM-based methods may serve as a first approximation for exploring the transitions between not-too-distant pairs of functional states. In the Chapter 3.0 , we will present a new approach based on the ANM [20, 21] to this aim, and apply it to the allosteric transitions of the GroEL-GroES complex (~ 8,000 residues, or ~ 1 MDa). The approach, referred to as *a*ANM, utilizes the ANM modes to guide the motion of the complex along the directions intrinsically favored by its instantaneous inter-residue contact topology, starting from both ends. The process ends when a common (intermediate) structure is reached, which is considered as the putative transition state [103].

2.0 VIBRATIONAL DYNAMICS OF ICOSAHEDRALLY SYMMETRIC BIOMOLECULAR ASSEMBLIES[1]

2.1 INTRODUCTION

Recent years have seen growing numbers of studies that resort to NMA as a simple, yet physically meaningful, means of studying the dynamics of proteins under equilibrium conditions [15]. NMA has been a method used in computational biology for more than two decades [104-106]. While NMA is much more efficient than methods based on full atomic simulations, it becomes prohibitively expensive as the size of the biomolecular system increases, due to the computational cost of energy minimization at the atomic scale, as well as the eigenvalue decomposition of increasingly larger Hessian matrices.

In the present study we explore the utility of carrying coarse-graining to its extreme by substituting continuum models for discretized ones [107]. A major utility of adopting a continuum model is to be able to derive analytical solutions that can assist in assessing the basic aspects of the complex processes being investigated. Here, a continuous mass density replaces the discrete masses, elastic moduli (Young's modulus and shear modulus) replace the springs linking the masses, and continuum elastic wave equations replace the usual Newton's equations

[1] The content of this chapter has been published on *Biophys. J.* 2009 [107].

of motion. While the discrete Newton's equations yield a set of coupled ordinary differential equations that can be solved using matrix diagonalization, the continuum wave equation is a partial differential equation. Provided the material is spatially isotropic and homogeneous, and the boundary conditions not too complicated, the wave equation can be solved analytically.

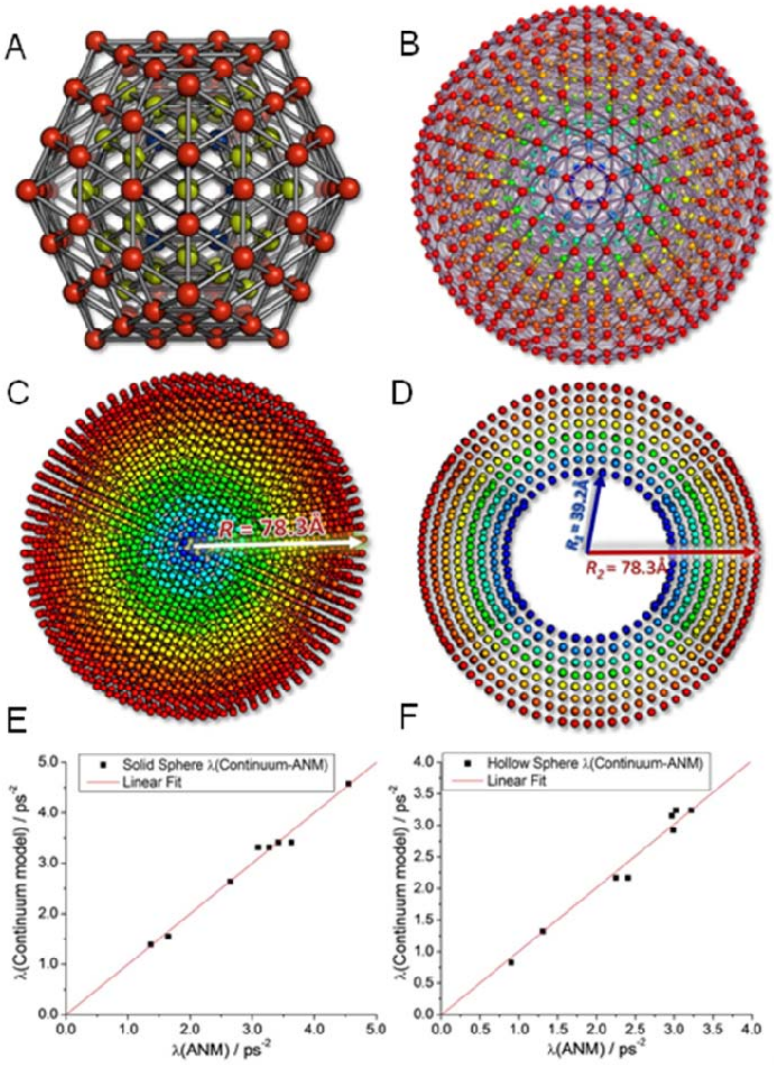

Figure 2. Mackay icosahedrons and corresponding toy models of spherical symmetry.

(A) Three-layer Mackay icosahedrons. **(B)** Sphere constructed using Mackay icosahedral symmetry for the nodes. **(C)** Cross-sectional view of the sphere displayed in panel b. **(D)** Cross-sectional view of a hollow sphere with indicated inner and outer radii. Panels (a-d) are colored by layer radius, from blue (small) to red (large). **(E)**

Comparison of the dispersion of modes (distribution of eigenvalues) predicted by ANM and by the continuum model for the model shown in panel **(C)**. **(F)** Same as panel **(E)**, for the hollow sphere model in **(D)**.

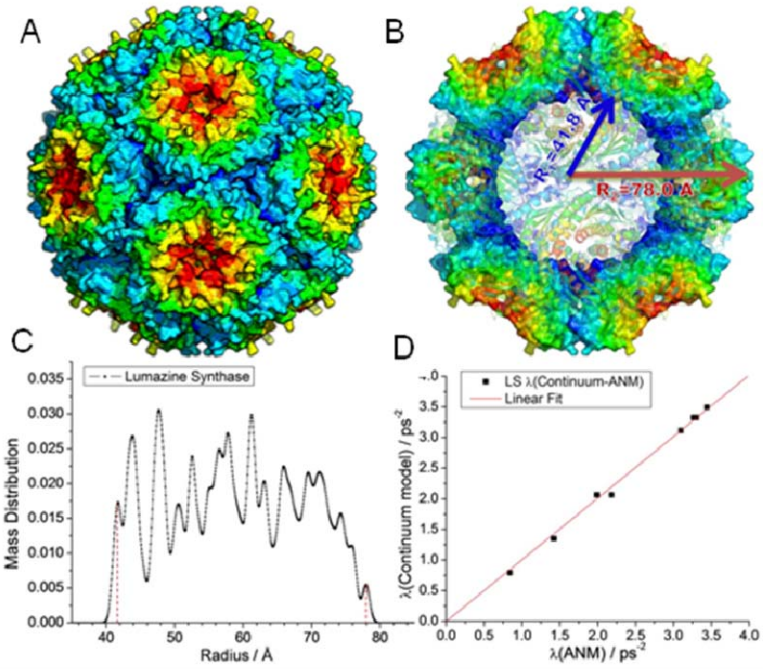

Figure 3. Results for lumazine synthase (LS).

(A) 3D structure of LS (PDB file: 1NQW [41]), with the surface colored by geometric position to illustrate the icosahedral symmetry. **(B)** Cross-sectional view of the same structure. **(C)** mass distribution as a function of radial position, used as basis for defining the inner and outer radii of the hollow sphere model used in continuum elasticity theory, and **(D)** comparison of the ANM eigenvalues with those found by the continuum elasticity theory for a hollow sphere with the same dimensions and packing density.

We solve the continuum elastic wave equations for systems with spherical symmetry, exploiting the symmetry to express our solutions using the natural basis set provided by vector spherical harmonics. We compare the collective dynamics predicted for biomolecular systems using discretized models with the vibrational spectra of solid and hollow spheres (see for example Figure 1). Specifically, the analytical solutions based on the continuum elastic wave

equations are compared with those obtained with simple toy models based on the Mackay icosahedron (see Figure 2) [108] and with naturally occurring supramolecular assemblies: lumazine synthase (LS, see Figure 3) [41], an enzyme responsible for the synthesis of riboflavin (vitamin B_2); satellite tobacco mosaic virus (STMV, see Figure 4 and Figure 5) [109], one of the smallest viruses known; and a triangulation number $T = 1$ particle of brome mosaic virus (BMV, see Figure 6) [110], a virus that infects a type of grass known as bromus.

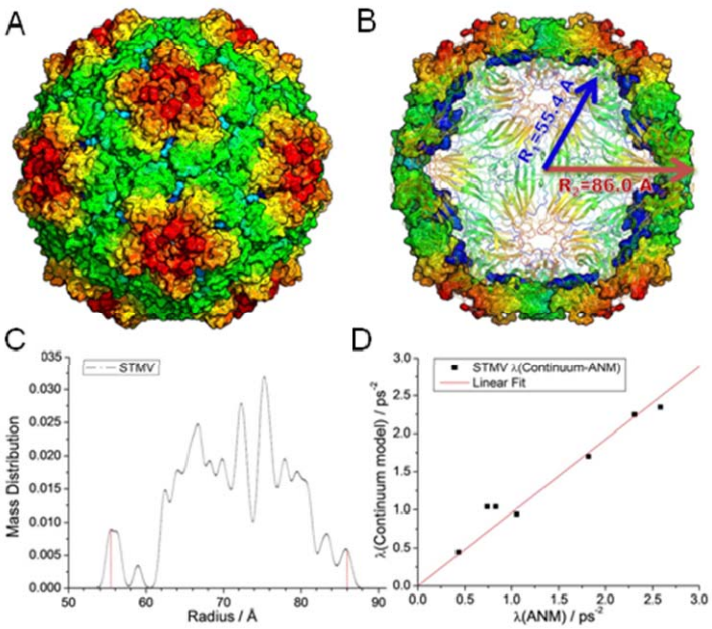

Figure 4. Results for STMV capsid.

(A) 3D structure of STMV (PDB file: 1A34 [109], capsid only), with the surface colored by geometric position to illustrate the icosahedral symmetry. **(B)** Cross-sectional view of the same structure. **(C)** mass distribution as a function of radial position, used as basis for defining the inner and outer radii of the hollow sphere model used in continuum elasticity theory, and **(D)** comparison of the ANM eigenvalues with those found by the continuum elasticity theory for a hollow sphere with the same dimensions and packing density.

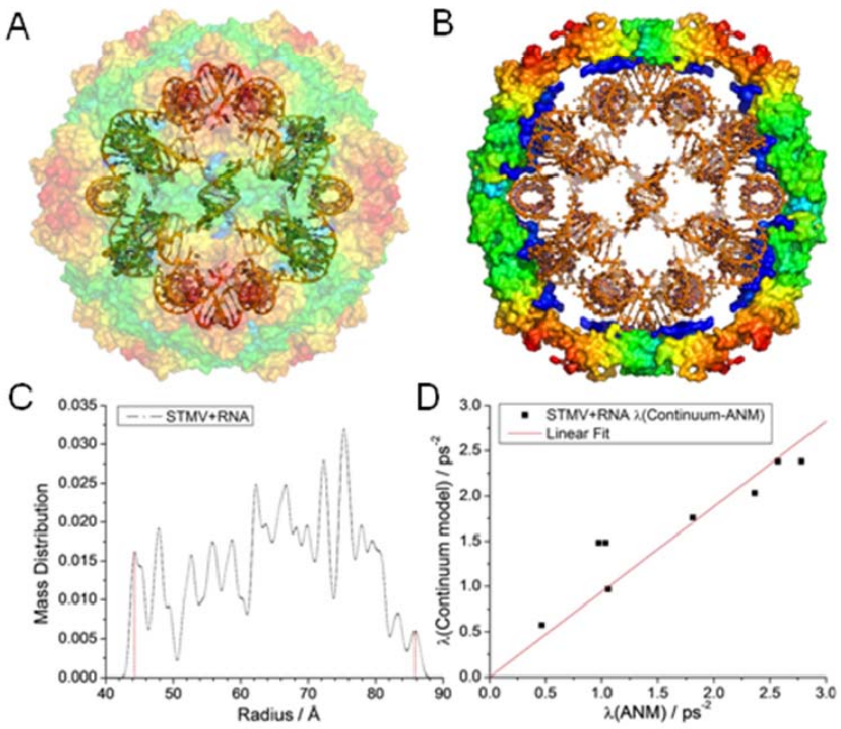

Figure 5. Results for STMV.

(A) 3D structure of STMV (PDB file: 1A34 [109], capsid plus genome), with the surface colored by geometric position to illustrate the icosahedral symmetry. **(B)** Cross-sectional view of the same structure. **(C)** mass distribution as a function of radial position, used as basis for defining the inner and outer radii of the hollow sphere model used in continuum elasticity theory, and **(D)** comparison of the ANM eigenvalues with those found by the continuum elasticity theory for a hollow sphere with the same dimensions and packing density.

Our calculations show that the low-lying vibrational frequencies and their associated normal modes closely follow patterns predicted by spherical symmetry and can be reproduced using a small number of parameters defined by the elasticity theory. Deviations from these predictions can be interpreted as weak perturbations that lower the spherical symmetry to icosahedral symmetry, which is naturally selected by many biological molecules. Comparison of discretized models with continuum theory of elasticity predictions permit us to assess the

macroscopic mechanical properties of the examined biomolecular systems, such as their Young's moduli (normalized with respect to the ANM force constants) and their Poisson's ratios.

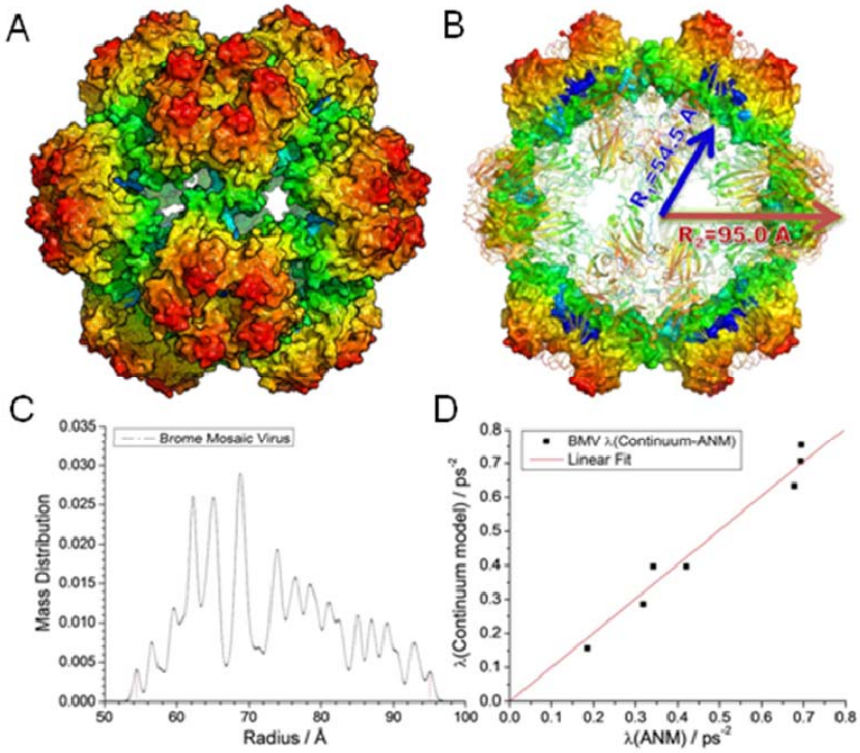

Figure 6. Results for BMV.

(A) 3D structure of BMV (PDB file: 1YC6 [110]), with the surface colored by geometric position to illustrate the icosahedral symmetry. **(B)** Cross-sectional view of the same structure. **(C)** mass distribution as a function of radial position, used as basis for defining the inner and outer radii of the hollow sphere model used in continuum elasticity theory, and **(D)** comparison of the ANM eigenvalues with those found by the continuum elasticity theory for a hollow sphere with the same dimensions and packing density.

2.2 THEORY

2.2.1 Continuum theory of elasticity

For an isotropic elastic medium, the equation of motion can be written as a wave equation [111]

$$\rho\ddot{\mathbf{u}} = \mu\nabla^2\mathbf{u} + (\lambda+\mu)\nabla\nabla\cdot\mathbf{u}, \tag{2.1}$$

where μ is the shear modulus, λ is the Lamé coefficient, and ρ is the density of the material, $\mathbf{u} = \mathbf{u}(\mathbf{r},t)$ represents the displacement from equilibrium and ∇ is the gradient operator with respect to the position vector $\mathbf{r}$, given by the coordinates. The double dot designates the second derivative with respect to time (i.e., acceleration in the present case), and $\nabla\cdot\mathbf{u}$ is the divergence of $\mathbf{u}$. Specifying the elastic constants λ and μ is equivalent to specifying the Young's modulus E and Poisson's ratio σ, in particular [111]

$$\mu = \frac{E}{2(1+\sigma)}, \quad \lambda+\mu = \frac{E}{2(1+\sigma)(1-2\sigma)}. \tag{2.2}$$

The Poisson ratio is defined as the ratio of the strains along the transverse and axial directions in response to an axial stress. In the case of presently considered spherical models, these two respective directions are, for example, the tangential and radial directions.

In a normal mode, the entire medium/system oscillates at a single temporal frequency ω. The solution then separates into spatial and temporal parts, so that $\mathbf{u}$ can be written as

$$\mathbf{u} = \mathrm{Re}\left(\mathbf{u}_0(\mathbf{r})e^{-i\omega t}\right). \tag{2.3}$$

If the divergence of $\mathbf{u}_0$ is equal to zero (i.e. $\nabla\cdot\mathbf{u}_0 = 0$), then substitution of Eq 2.3 into the wave equation (Eq. (2.1)) yields

$$\nabla^2 \mathbf{u}_0 + k^2 \mathbf{u}_0 = 0 \tag{2.4}$$

where $k = \omega / c_t$ is recognized as a *transverse* wavenumber, and the transverse sound speed is given by

$$c_t = \sqrt{\frac{\mu}{\rho}} = \sqrt{\frac{E}{2(1+\sigma)\rho}}. \tag{2.5}$$

If, instead, $\mathbf{u}_0$ is curl-free (i.e. $\nabla \times \mathbf{u}_0 = 0$), then, using $\nabla\nabla \cdot \mathbf{u}_0 = \nabla^2 \mathbf{u}_0 + \nabla \times \left(\nabla \times \mathbf{u}_0 \right)$, the wave equation reduces to

$$\nabla^2 \mathbf{u}_0 + q^2 \mathbf{u}_0 = 0 \tag{2.6}$$

where $q = \omega / c_l$ is the *longitudinal* wavenumber, and c_l is the longitudinal sound speed given by

$$c_l = \sqrt{\frac{\lambda + 2\mu}{\rho}} = \sqrt{\frac{E(1-\sigma)}{\rho(1+\sigma)(1-2\sigma)}}. \tag{2.7}$$

Note that the longitudinal and transverse wavenumbers obey the relation

$$\frac{q}{k} = \frac{c_t}{c_l} = \sqrt{\frac{1-2\sigma}{2(1-\sigma)}}. \tag{2.8}$$

When analyzing problems with spherical symmetry it is advantageous to represent the displacement field $\mathbf{u}_0$ on the basis of *spherical harmonics* and *spherical Bessel functions*. Solutions to Eq. (2.1) take the generic form [112-115]

$$\mathbf{u}_0(\boldsymbol{r}) = c_0 \nabla \varphi(\boldsymbol{r}) + c_1 \mathbf{L} \psi(\boldsymbol{r}) + c_2 \nabla \times \mathbf{L} \psi(\boldsymbol{r}) \tag{2.9}$$

where $\mathbf{L} = \boldsymbol{r} \times \nabla$ is the angular momentum operator, and

$$\begin{aligned} \varphi(\boldsymbol{r}) &= j_l(qr) Y_{lm}(\theta, \phi) \\ \psi(\boldsymbol{r}) &= j_l(kr) Y_{lm}(\theta, \phi). \end{aligned} \tag{2.10}$$

Here $Y_{lm}(\theta,\phi)$ is a spherical harmonic, j_l is the spherical Bessel function of the *first kind* and is analytic at the spatial coordinate $\mathbf{r} = 0$. Vector fields of these forms, known as *vector spherical harmonics*, are discussed in more detail in the next section. If the sphere is hollow, so that our solution does not include the point $\mathbf{r} = 0$, then we add to $\mathbf{u}_0$ a function of the form

$$\tilde{\mathbf{u}}_0(\boldsymbol{r}) = d_0 \nabla \tilde{\varphi}(\boldsymbol{r}) + d_1 \mathbf{L}\tilde{\psi}(\boldsymbol{r}) + d_2 \nabla \times \mathbf{L}\tilde{\psi}(\boldsymbol{r}) \tag{2.11}$$

where

$$\begin{aligned} \tilde{\varphi}(\boldsymbol{r}) &= n_l(qr)Y_{lm}(\theta,\phi) \\ \tilde{\psi}(\boldsymbol{r}) &= n_l(kr)Y_{lm}(\theta,\phi). \end{aligned} \tag{2.12}$$

Here n_l is the spherical Bessel function of the *second kind* and is nonanalytic at the origin.

The three terms in Eq. (2.9) (and also in Eq. (2.11)) have physical interpretations related to sound waves. The first term, as the gradient of a scalar function, is longitudinal in the sense that the direction of the vector $\mathbf{u}_0$ is parallel to the direction in which the amplitude varies. The second term is transverse because it is perpendicular to the first. It is also perpendicular to the radial direction, that is at every point ***r*** it lies tangent to the sphere of radius *r*. The third term is also transverse, and is perpendicular to the first two terms.

Discrete allowed wavenumbers are determined by the boundary conditions. For free boundaries, the appropriate condition is vanishing normal component of the stress tensor $\boldsymbol{\tau}$. Elements of $\boldsymbol{\tau}$ can be written in term of the elements of strain tensor, u_{ik}, which can be derived from the displacement vector, $\mathbf{u}_0$, as [111]

$$\tau_{ik} = \frac{E}{1+\sigma}\left(u_{ik} + \frac{\sigma}{1-2\sigma}u_{ll}\delta_{ik}\right), \tag{2.13}$$

where u_{ll} is the trace of the strain tensor, and δ_{ik} is the Kronecker delta. The normal component vanishes if $\boldsymbol{\tau} \cdot \hat{\mathbf{n}} = 0$, where $\hat{\mathbf{n}}$ is the normal to the surface at the boundary. In spherical coordinates i and k refer to the unit vectors $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$ or $\hat{\boldsymbol{\phi}}$. The appropriate boundary conditions for a sphere of radius R are therefore [112]

$$\boldsymbol{\tau}_{i\hat{\boldsymbol{r}}}\big|_{r=R} = 0 . \tag{2.14}$$

Since i takes on three allowed values (corresponding to the curvilinear directions $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$), Eq. (2.14) constitutes a set of three linearly independent equations. Similarly, for a hollow sphere we obtain two sets of equations [113]

$$\boldsymbol{\tau}_{i\hat{\boldsymbol{r}}}\big|_{r=R_1} = 0 \quad \text{and} \quad \boldsymbol{\tau}_{i\hat{\boldsymbol{r}}}\big|_{r=R_2} = 0 . \tag{2.15}$$

where R_1 and R_2 are the inner and outer radii, respectively, yielding a set of six linearly independent equations.

The equations subject to these boundary conditions are solved using methods of linear algebra as discussed in Appendix B. Solutions exist only for well-defined quantized values of the wavenumbers q and k, and therefore only for special vibrational frequencies $\omega = c_l q = c_t k$.

Each mode has a degeneracy arising from spherical symmetry. Specifically, the frequency depends on the total angular momentum index l but does not depend on the azimuthal angular momentum index m, which range from $-l$ to $+l$ by integers. Therefore each allowed frequency has degeneracy

$$\Omega_l = 2l + 1 = 1, 3, 5, 7, \ldots . \tag{2.16}$$

2.2.2 Spherical harmonics

Spherical harmonics and vector spherical harmonics, respectively, comprise complete basis sets for scalar and and vector fields on the sphere. Consider a scalar field $\rho(\theta,\phi)$. Completeness enables us to write

$$\rho(\theta,\phi) = \sum_{l,m} \rho_{lm} Y_{lm}, \tag{2.17}$$

and we can use orthogonality of the spherical harmonics to obtain the expansion coefficients

$$\rho_{lm} = \int Y^{\star}_{lm}(\theta,\phi)\rho(\theta,\phi)d\Omega, \tag{2.18}$$

where $d\Omega = \sin\theta d\theta d\phi$ ranges over 4π of solid angle. Similarly, an arbitrary vector field $\mathbf{u}(\theta,\phi)$ can be expressed in terms of vector spherical harmonics [116] as

$$\mathbf{u}(\theta,\phi) = \sum_{l,m} \left[A_{lm}\mathbf{V}_{lm}(\theta,\phi) + B_{lm}\mathbf{W}_{lm}(\theta,\phi) + C_{lm}\mathbf{X}_{lm}(\theta,\phi) \right] \tag{2.19}$$

The expansion coefficients in the above equation can be obtained through

$$A_{lm} = \int \mathbf{V}^{\star}_{lm}(\theta,\phi) \bullet \mathbf{u}(\theta,\phi)\, d\Omega \tag{2.20}$$

and analogous equations hold for B_{lm} and C_{lm}.

Projection onto spherical harmonics provides a simple way to classify the symmetry of scalar functions. If $\rho(\theta,\phi)$ has spherical symmetry (i.e., is a constant, independent of θ and ϕ) then its projection onto Y_{lm} vanishes except for $l = m = 0$. If instead $\rho(\theta,\phi)$ has icosahedral symmetry (i.e., is not constant but is invariant under rotations belonging the the icosahedral symmetry group) then its projection onto Y_{lm} vanishes except for $l = 0,6,10,12,...$ [117]. If

$\rho(\theta,\phi)$ has nonvanishing projections onto other values of l then it has symmetry lower than icosahedral.

For a given l the coefficients ρ_{lm} depend on the choice of spatial coordinates. We can form a rotationally invariant measure of the angular momentum projections as

$$\rho_l = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^{l} | \rho_{lm} |^2 \right]^{1/2} . \tag{2.21}$$

Then ρ_6 / ρ_0 measures the reduction of spherical symmetry to icosahedral, and the ratio of ρ_l / ρ_6 measures the degree to which icosahedral symmetry is broken. For example, nonzero ρ_4 indicates tetrahedral or cubic symmetry breaking.

Alternatively, we can express icosahedrally symmetric functions of spatial position as a superposition of "icosahedral harmonics" [118]. The simplest icosahedral harmonic I_0 is just a constant. The next icosahedral harmonic, I_6, is a specific combination of Y_{6m} spherical harmonics that remains invariant under all rotations that are symmetries of an icosahedron. After that comes I_{10}, etc. A combination such as $c_0 I_0 + c_6 I_6$ with $c_6 << c_0$ represents nearly perfect spherical symmetry with a weak icosahedral symmetry breaking.

The vector spherical harmonics $\mathbf{V}_{lm}$, $\mathbf{W}_{lm}$ and $\mathbf{X}_{lm}$ [116] are only one of many choices for the vector field basis set. In our analysis we also use the alternate set in Eq. (2.9). Given the scalar function $\phi(\mathbf{r}) = F(r) Y_{lm}(\theta,\phi)$, its gradient and angular momentum can be expressed as linear combinations of the vector spherical harmonics through

$$\nabla\phi(\boldsymbol{r}) = \left(\frac{l}{r}F - \frac{dF}{dr}\right)\left(\frac{l+1}{2l+1}\right)^{1/2}\mathbf{V}_{lm} + \left(\frac{l+1}{r}F + \frac{dF}{dr}\right)\left(\frac{l}{2l+1}\right)^{1/2}\mathbf{W}_{lm}$$
$$\mathbf{L}\phi(\boldsymbol{r}) = i\left(l(l+1)\right)^{1/2}F\mathbf{X}_{lm} \tag{2.22}$$
$$\nabla\times\left(\mathbf{L}\phi(\boldsymbol{r})\right) = l\left(\frac{l}{r}F - \frac{dF}{dr}\right)\left(\frac{l+1}{2l+1}\right)^{1/2}\mathbf{V}_{lm} - (l+1)\left(\frac{l+1}{r}F + \frac{dF}{dr}\right)\left(\frac{l}{2l+1}\right)^{1/2}\mathbf{W}_{lm}$$

Notice that the vector spherical harmonics of type X_l are always tangent to the sphere. We refer to these modes as "torsional". In contrast vector spherical harmonics of types V_l and W_l contain both tangent and radial components, and usually enter solutions of the wave equation in combination with each other. We refer to these modes as "spheroidal".

2.2.3 Spherical, icosahedral and tetrahedral symmetry

The solutions to the equation of motion Eq. (2.1) are constrained by the requirements of spherical symmetry [119, 120]. Given a solution with a specific vibrational frequency (i.e. a normal mode), any spatial rotation generates a new solution of identical vibrational frequency. The set of all solutions of identical frequency form an invariant set known as an *irreducible representation*. A linearly independent subset forms a *basis* of the irreducible representation. The number of elements in the basis is known as the *dimension* of the irreducible representation and corresponds to the degeneracy of the vibrational mode.

Spherical harmonics provide basis sets for irreducible representations of the spherical rotation symmetry group $SO(3)$. For a given angular momentum l the set $\{Y_{lm}, m = -l,...,+l\}$ constitutes a basis for an Ω_l-dimensional irreducible representation. Consequently any normal mode $\mathbf{u}_0$ can be expressed in terms of spherical harmonics multiplying a function of radius r, as in Eqs. (2.10) and (2.12), or equivalently using the vector spherical harmonics. Conversely,

given a normal mode $\mathbf{u}_0$ obtained from the diagonalization of the Hessian matrix, we may identify its angular momentum content by projecting it onto vector spherical harmonics as shown in section 2.2.2.

At the atomistic level, biological structures always break spherical symmetry, with icosahedral rotation symmetry being prevalent [121]. Owing to the chirality of the protein subunits, we consider only pure rotations, without inversion or reflection [122-124]. The symmetry-breaking may be confined to the internal structure, with the external boundaries remaining nearly spherical, as in the case of LS and STMV. In other cases, the external shape may strongly break the spherical symmetry as in the case of the $T = 1$ particle of BMV. This distinction is illustrated by a "toy" model we introduce based on the Mackay Icosahedron [108]. The Mackay icosahedron consists of 20 independent grains of face-centered cubic crystal arranged with special orientations inside an icosahedron (See Figure 2 A). By definition, it exhibits both internal and external icosahedral symmetry. We can utilize this structure to generate a spherical object where the nodes are icosahedrally arranged. To this aim, we project each node along the radial direction, so that they are located at the same radial distance, forming concentric shells (see Figure 2 B). By varying the inner and outer radius, one can obtain corresponding solid or hollow spheres (See Figure 2 C and D). For example, we create an analogue of the structure of LS, by setting the inner and outer radii of the hollow sphere in a 2 : 1 ratio.

Because the symmetry group Y of the icosahedron comprises a finite subset of the full rotational symmetry group of a sphere, the icosahedral symmetry group has finitely many irreducible representations. As listed in Table 1, these are: the one-dimensional unit representation A ; two three-dimensional representations F_1 and F_2 ; a four dimensional

representation G; a five-dimensional representation H [120]. Since the dimensionality of the irreducible representations determines the degeneracy of each normal mode, we see that the allowed degeneracies of vibrational frequency are 1 (nondegenerate), 3, 4 and 5. No other degeneracies may occur. This pattern of degeneracies differs from the spherical case as given in Eq. (2.16).

Table 1. Icosahedral group character showing angular momenta *l* and irreducible representations ("irreps").

l	irreps	$1C_0$	$15C_2$	$20C_3$	$12C_5$	$12C_5^2$
0	A	1	1	1	1	1
1	F_1	3	-1	0	τ*	$-\tau^{-1}$
2	H	5	1	-1	0	0
3	F_2	3	-1	0	$-\tau^{-1}$	τ
3	G	4	0	1	-1	-1
4	$G + H$	9	1	0	-1	-1
5	$F_1 + F_2 + H$	11	-1	-1	1	1
6	$A + F_1 + G + H$	13	1	1	τ	$-\tau^{-1}$

* $\tau = (\sqrt{5}+1)/2$ is the Golden Mean, and C_n indicates an axis of n-fold rotational symmetry.

The subgroup relationship between the icosahedral group and the full rotation group, $Y \subset SO(3)$, allows us to relate irreducible representations of Y to the spherical harmonics Y_{lm}. We identify $l = 0$ with A, $l = 1$ with F_1 and $l = 2$ with H. For $l = 3$ and beyond, the spherical harmonics contain multiple irreducible representations of Y, as listed in Table 1. For instance, $l = 3$ contains both F_2 and G. The $l = 0$ spherical harmonic, Y_{00}, is the unit representation of the spherical symmetry group, while the unit representation A of the icosahedral group is contained in $l = 0, 6, 10, 12, \ldots$. Notice this pattern reproduces the pattern of projections of icosahedrally symmetric structures discussed in section 2.2.2.

We will occasionally need to consider a further reduction of symmetry, to that of a tetrahedron with symmetry group [119] $T \subset Y$. The tetrahedral group T has irreducible representations A (dimension 1), E (dimension 2) and F (dimension 3). The allowed degeneracies of vibrational frequencies are thus 1, 2 and 3. Irreducible representations of Y are sometimes reducible as representations of T . We can write $A = A$, $F_1 = F$, $F_2 = F$, $G = A + F$, and $H = E + F$, where the left-hand side of the equation corresponds to icosahedral representations and the right-hand side corresponds to tetrahedral representations.

For a given vibrational mode (i.e., eigenvector of the Hessian $\mathbf{H}$) we can form arbitrary linear combinations of degenerate modes and the result is still a mode (eigenvector) of the same frequency. When considering modes of a given type of vector spherical harmonic, we take linear combinations in order to make them real rather than complex, and we group them further into combinations corresponding to irreducible representations of the icosahedral and tetrahedral symmetry groups.

2.2.4 Calculational procedures

In the following, we compare the eigenvalue spectra obtained from ANM with the predictions of continuum elasticity theory. We do not know *a priori* the elastic constants E and σ (or equivalently μ and λ , or sound speeds c_t and c_l), so we determine these by least squares fitting to the eigenvalues obtained from ANM. Elastic constants E , λ and μ have units of Pa (or $\mathrm{N/m^2}$), while the Poisson's ratio σ is dimensionless. The ANM force constant γ has units of N/m (note the conversion factor 1 N/m = 1.44 kcal/(mol $\cdot$ Å^2)). Eigenvalues $\lambda = \omega^2$ will be quoted in units of $(\mathrm{rad/ps})^2$. The optimal fitting of the continuum elastic constants to ANM

eigenvalues yield the relative Young's modulus E/γ which we quote in units of nm^{-1}. The uncertainty in the relative Young's modulus is also estimated from the linear fit.

To construct the ANM, we take a cutoff distance of r_c = 15Å, and we assign an average amino acid mass of 108 Da to each node (in the case of STMV+RNA, the node mass is 120 Da.) The hollow sphere continuum models are defined in a consistent manner by using the first and last peaks in the mass distribution functions as the inner and outer radii R_1 and R_2, respectively (see Figure 4 and Figure 5). For the hollow sphere model representing the LS structure, we take $R_1 = 41.8$ Å and $R_2 = 78.0$ Å. The mass density is obtained from $\rho = M/V$, where M is the total mass and V the hollow sphere volume.

Using the solvability conditions (Eqs. (A.4) and (A.5) in the Appendix A), we calculate the allowed wavenumbers k and q arising from boundary conditions for a given geometry (Eqs. (2.14) or (2.15)). Note that k and q depend on the Poisson's ratio σ but not the Young's modulus E or the density ρ. Then, from the sound speeds (Eqs. (2.5) and (2.7)) we predict the resonance frequency $\omega = c_t k = c_l q$.

To fit the elastic constants, we define a mean-square difference (MSD) between the ANM and the continuum eigenvalues,

$$\mathrm{MSD} = \frac{1}{N}\sum_n \left(\lambda_n^{(ANM)} - \lambda_n^{(cont)}\right)^2, \tag{2.23}$$

where the sum runs over each distinct eigenvalue λ_n whose eigenvector $\mathbf{v}^{(n)}$ has angular momentum $l \le 3$. For a given σ, the optimal Young's modulus E can be determined analytically by minimizing the MSD using the relation,

$$\lambda^{(cont)} \equiv c_t^2 k^2 = \frac{E}{2\rho(1+\sigma)} k^2 . \tag{2.24}$$

In this manner, we obtain the MSD for arbitrary σ, then select the optimal σ that minimizes the MSD.

2.3 RESULTS

2.3.1 Overview

The results obtained for the four biological structures analyzed in the present study, lumazine synthase (LS), satellite tobacco mosaic virus (STMV) empty capsid, STMV containing its genetic material (RNA), and $T = 1$ particle of brome mosaic virus (BMV), are presented in the respective Figure 3, Figure 4, Figure 5 and Figure 6. In each case, the panel **(A)** displays the molecular system color-coded by the geometric position (see below), panel **(B)** a cross-sectional view, panel **(C)** the mass distribution as a function of radial position, and panel **(D)** the comparison of the dispersions of modes predicted by the ANM and the continuum model. A principal result from the present study is the identification of the types of modes operating in each system, expressed in terms of their icosahedral and vector spherical representations, as summarized Table 2 and Table 4 for toy models and in Table 5-Table 8 for biological systems (see also Figure 1). Table 3 summarizes the results, i.e., the Poisson's ratios and effective Young's moduli corresponding to each system (columns 3 and 4), and the correlation between the two sets of data in terms of MSD (Eq. (2.23)) and correlation coefficient (columns 5 and 6). More details for each examined system will be presented below.

Table 2. ANM eigenvalues and mode types for solid sphere toy model.

Mode #	Eigenvalue (ps^{-2})	Icosahedral Representations	Vector Spherical Representations
1-5	1.36451	H	$\mathbf{X}_2$
6-10	1.64639	H	$\mathbf{V}_2 + \mathbf{W}_2$
11-13	2.64211	F_1	$\mathbf{V}_1$
14-17	3.09554	G	$\mathbf{X}_3$
18-20	3.26970	F_2	
21-23	3.42060	F_2	$\mathbf{V}_3 + \mathbf{W}_3$
24-27	3.63187	G	
28	4.55380	A	$\mathbf{V}_0$

2.3.2 Results for toy models

Before we proceed to the results for the four biomolecular systems, we first note that the results for the filled and hollow spheres constructed from the Mackay icosahedron confirm that the mode spectra from ANM and continuum elasticity model yield a remarkably high correlation (respective MSD values of 0.0136 ps^{-4} and 0.0192 ps^{-4}, and correlation coefficients above 0.99; see Figure 2 and Table 3) in so far as the distribution of eigenvalues are concerned. The comparison yields Poisson's ratios of σ = 0.26 and 0.27, and relative Young's moduli E/γ = 31.2 ± 0.4 nm^{-1} and 32.8 ± 0.7 nm^{-1}, respectively. As discussed in Section 2.5, spherically symmetric modes of high angular momentum (e.g. $l = 3$) split into multiple icosahedrally symmetric modes (e.g. F_2 and G) with different eigenvalues. For our toy models, the reduction of spherical to icosahedral symmetry is weak, so the eigenvalue splitting is small, as is evident in Table 2 and Table 4 (e.g. compare eigenvalues of F_2 and G).

Table 3. Results from fitting the eigenvalues to continuum models.

	density ρ (kg/m^3)	Poisson's Ratio σ	Relative Young's Modulus E/γ (nm^{-1})	Mean-squared difference (ps^{-4})	Fitting correlation
Solid Sphere	905.99	0.26	31.2 ± 0.4	0.0136	0.9984
Hollow Sphere	941.53	0.27	32.8 ± 0.7	0.0192	0.9966
Lumazine Synthase	987.60	0.30	36.2 ± 0.3	0.0040	0.9993
STMV	823.82	0.24	26.0 ± 1.1	0.0317	0.9863
STMV+RNA	981.79	0.20	28.5 ± 1.7	0.0980	0.9674
BMV	561.39	0.30	6.58 ± 0.2	0.0016	0.9932

Table 4. ANM eigenvalues and mode types for hollow sphere toy models.

Mode #	**Eigenvalue (ps^{-2})**	**Icosahedral Representations**	**Vector Spherical Representations**
1-5	0.89990	H	$\mathbf{V}_2 + \mathbf{W}_2$
6-10	1.30453	H	$\mathbf{X}_2$
11-13	2.24883	F_2	$\mathbf{V}_3 + \mathbf{W}_3$
14-17	2.40630	G	
18-20	2.96895	F_1	$\mathbf{V}_1$
21	2.98536	A	$\mathbf{V}_0$
22-25	3.02320	G	$\mathbf{X}_3$
26-28	3.21800	F_2	

2.3.3 Lumazine synthase

Lumazine synthase (LS) is a hollow, nearly spherical molecule consisting of 60 identical subunits arranged with icosahedral symmetry. ANM calculations were based on the crystallographic structure resolved in a body-centered cubic lattice with space group I23, deposited in the PDB (ID:1NQW [41]). As a result, the structure exhibits tetrahedral symmetry, which may be viewed as a reduced form of icosahedral symmetry. Accordingly, the degeneracies of the accessible vibrational modes are 1, 2 and 3 (see Table 5). For example, the 5-fold degenerate (lowest frequency) mode observed in the case of the solid and hollow spheres is now

split into two sets of modes, with degeneracies 2 and 3. Note that the distinction between the frequencies of these two sets of modes appears only at the third significant digits of the corresponding eigenvalues.

Table 5. ANM eigenvalues and mode types for lumazine synthase.

Mode #	Eigenvalue (ps^{-2})	Tetrahedral Representations	Icosahedral Representations	Vector Spherical Representations
1-2	0.83817	*E*	*H*	$\mathbf{V}_2 + \mathbf{W}_2$
3-5	0.84045	*F*		
6-8	1.41542	*F*	*H*	$\mathbf{X}_2$
9-10	1.42215	*E*		
11-13	1.98688	*F*	*G*	$\mathbf{V}_3 + \mathbf{W}_3$
14	1.99022	*A*		
15-17	2.18280	*F*	F_2	
18	3.09788	*A*	*A*	$\mathbf{V}_0$
19-21	3.26416	*F*	F_2	$\mathbf{X}_3$
22-24	3.29502	*F*	*G*	
25	3.30125	*A*		
26-28	3.44217	*F*	F_1	$\mathbf{V}_1$

The continuum elasticity model adopted for LS requires the definition of the inner and outer radii of the representative sphere, which were deduced from the mass distribution calculated as a function of radial distance (panel C of Figure 3). The red dotted bars delimit therein the inner and outer radii as $R_1 = 41.8$ Å and $R_2 = 78.0$ Å, respectively. Table 5 lists the representations of the ANM modes in the icosahedral and spherical symmetry groups. For example, the lowest frequency (doubly degenerate) mode (modes 1 and 2) corresponds to the vector spherical harmonics $\mathbf{V}_2$ and $\mathbf{W}_2$ (see section 2.2.2), while the non-degenerate mode 18 corresponds to $\mathbf{V}_0$, as illustrated in Figure 1. The comparison of the outputs from the two methods permit us to evaluate the two macroscopic quantities $E/\gamma = 36.2 \pm 0.3\ \mathrm{nm}^{-1}$ and $\sigma =$

0.30 (with MSD of 0.0040 ps^{-4}) for the respective relative Young's modulus and Poisson's ratio that best represent the global relaxational behavior of LS.

2.3.4 Satellite tobacco mosaic virus capsid

The ANM calculations for STMV were performed using the PDB structure (ID:1A34 [109]). Note that STMV has triangulation value of $T = 1$, i.e., it is composed of 60 icosahedrally arranged identical subunits. In this case, the reduction of spherical symmetry to icosahedral is slightly stronger (see Figure 4), however no breaking of icosahedral symmetry due to crystallographic lattice type is evident. The eigenvalues calculated with ANM are listed in Table 6 for the empty capsid, and in Table 7 for the capsid with RNA. Taking R_1 = 55.4 Å for the empty capsid and R_2 = 86.0 Å , we find the eigenvalues match continuum elasticity theory with $E/\gamma = 26.0 \pm 1.1\ \mathrm{nm}^{-1}$ and $\sigma = 0.24$ with a MSD of 0.0317 ps^{-4}.

Table 6. ANM eigenvalues and mode types for satellite tobacco mosaic virus (STMV) capsid.

Mode #	Eigenvalue (ps^{-2})	Icosahedral Representations	Vector Spherical Representations
1-5	0.43242	H	$\mathbf{V}_2 + \mathbf{W}_2$
6-8	0.73901	F_2	$\mathbf{V}_3 + \mathbf{W}_3$
9-12	0.82876	G	
13-17	1.05426	H	$\mathbf{X}_2$
18-21	1.22444	G	$\mathbf{V}_4 + \mathbf{W}_4$
22-26	1.46058	H	
27	1.81925	A	$\mathbf{V}_0$
28-32	2.13790	H	$\mathbf{V}_5 + \mathbf{W}_5$
33-35	2.28670	F_2	
36-38	2.30968	F_1	$\mathbf{V}_1$

In addition to the empty STMV capsid, the same calculation was preformed for STMV partially filled with its genome [109]. For each RNA nucleotide, atoms P, C2 and C4' were taken as nodes of the ANM (see Figure 5). The inner radius R_1 drops from 55.4 Å to 44.2 Å. The resulting eigenvalues are listed in Table 7. Notice that the modes with $l = 1$ appear at a lower frequency when RNA is included ($l = 1$ appears in modes 28-30 with RNA; and modes 36-38 without RNA). Likewise, the eigenvalue of the $l = 2$ spheroidal squeezing modes (modes 1-5) increases by 8% in the presence of the genome, while the $l = 2$ torsional twisting modes (modes 13-17) increase by only 0.4%.

Table 7. ANM eigenvalues and mode types for intact STMV (capsid+RNA).

Mode #	Eigenvalue (ps^{-2})	Icosahedral Representations	Vector Spherical Representations
1-5	0.46662	H	$\mathbf{V}_2 + \mathbf{W}_2$
6-8	0.97430	F_2	$\mathbf{V}_3 + \mathbf{W}_3$
9-12	1.03662	G	
13-17	1.05818	H	$\mathbf{X}_2$
18-21	1.61848	G	$\mathbf{V}_4 + \mathbf{W}_4 + \mathbf{X}_3$
22	1.81089	A	$\mathbf{V}_0$
23-27	2.12701	H	$\mathbf{V}_4 + \mathbf{W}_4 + \mathbf{X}_2$
28-30	2.36294	F_1	$\mathbf{V}_1$

2.3.5 Brome mosaic virus capsid

The $T = 1$ particle of brome mosaic virus (ID:1YC6 [110], abbreviated as BMV) is assembled *in vitro* from the wild-type $T = 3$ brome mosaic virus under certain chemical conditions. Unlike STMV, it has a strong reduction of spherical symmetry to icosahedral ($\rho_6 / \rho_0 = 0.1029$, see section 2.2.2) and it proves useful to contrast the spherical continuum model with this extreme case.

The inner and outer radii are R_1 = 54.5 Å and R_2 = 95.0 Å, respectively. Despite the strong deviation from spherical symmetry, we obtain a reasonable fit between continuum elastic constants to our ANM model. We find E/γ = 6.58 ± 0.2 nm^{-1} and σ = 0.30 with MSD of 0.0016 ps^{-4} in this case. The lower magnitude of the Young's modulus for BMV compared with the other examined models is because the irregular structure of BMV (Figure 6) leads to an overall lower packing density.

Table 8. ANM eigenvalues and mode types for BMV.

Mode #	Eigenvalue (ps^{-2})	Icosahedral Representations	Vector Spherical Representations*
1-5	0.18428	H	$\mathbf{W}_2 + \mathbf{V}_2$ ($\mathbf{X}_2 + \mathbf{X}_4 + \mathbf{V}_4$)
6-10	0.31746	H	$\mathbf{X}_2$ ($\mathbf{X}_4 + \mathbf{W}_2 + \mathbf{V}_4$)
11-14	0.34180	G	$\mathbf{V}_3 + \mathbf{W}_3$ ($\mathbf{X}_3$)
15-17	0.42087	F_2	$\mathbf{V}_3 + \mathbf{W}_3$ ($\mathbf{X}_3$)
18	0.67744	A	$\mathbf{V}_0$ ($\mathbf{W}_6 + \mathbf{X}_6$)
19-22	0.69309	G	$\mathbf{X}_3$ ($\mathbf{V}_4 + \mathbf{V}_3 + \mathbf{W}_3$)
23-25	0.69450	F_1	$\mathbf{V}_1$ ($\mathbf{X}_5 + \mathbf{W}_5 + \mathbf{V}_5$)

* Vector spherical representations in parenthesis are secondary contributions.

Although we still identify a dominant vector spherical harmonic, as given in Table 8, we now find significant contributions from other vector spherical harmonics, as given in parentheses following the dominant contribution. These secondary contributions were identified on the basis of the projection method outlined in section 2.2.2. The secondary contributions follow patterns that can be explained on the basis of icosahedral symmetry together with addition of angular momentum laws that we present below. Although the ANM eigenvectors no longer correspond cleanly to vector spherical harmonics, we can still label them uniquely according to icosahedral irreducible representations.

For weakly icosahedral structures such as LS and STMV, each mode could be identified with a specific vector spherical harmonic type. In contrast, for the strongly icosahedral BMV structure we found admixtures of extra vector spherical harmonic types in each mode. This behavior is an expected result of icosahedral perturbations on an otherwise spherical structure. Specifically, each mode belongs to a unique icosahedral irreducible representation, but many different vector spherical harmonics may transform under the same icosahedral representation (e.g. according to Table 1 the nondegenerate mode of icosahedral type A belongs to angular momentum $l = 0$ and $l = 6$ but no angular momenta in between). The results for BMV obey the restrition of same icosahedral representation (e.g. according to Table 8, mode 18 of type A projects only onto $l = 0$ and $l = 6$).

The symmetry classification governs which combinations of modes are permissible, but does not guarantee they will all occur in practice. To understand what mixtures will actually occur requires a mechanism to couple the different vector spherical harmonics. Addition of angular momentum laws [119] provide one such mechanism. We start with the equation of motion, Eq. (2.1) in the form

$$\ddot{\mathbf{u}} = \frac{\mu}{\rho}\nabla^2\mathbf{u} + \frac{\lambda+\mu}{\rho}\nabla\nabla\cdot\mathbf{u} \tag{2.25}$$

and consider slow spatial variation of the material parameters μ/ρ and λ/ρ. These variations obey icosahedral symmetry and the parameters can therefore be expanded in a series of icosahedral harmonics (see section 2.2.2) as $c_0 I_0 + c_6 I_6$. When we substitute a vector spherical harmonic of angular momentum l for the variable u in the wave equation (Eq.(2.25)), it gets multiplied by a term of angular momentum $l' = 0$ and a relatively weak term of angular momentum $l' = 6$. According to laws of angular momentum addition [119], multiplication of l

by $l' = 0$ does not alter the value of l. However, multiplication of l by $l' = 6$ will produce a series of additional angular momenta l'', ranging from $6 - l$ up to $6 + l$.

In such a manner, modes of icosahedral type H and angular momentum $l = 2$ necessarily contain admixtures of momentum $l = 4$, and the nondegenerate mode of icosahedral type A and angular momentum $l = 0$ mixes with an icosahedrally symmetric combination of angular momentum $l = 6$. Angular momentum $l = 3$ largely remains intact. These patterns are visible in Table 8.

The mechanism just discussed preserves the torsional and spheroidal nature of the modes. Therefore modes should remain purely of type $\mathbf{X}$ or $\mathbf{V} + \mathbf{W}$. Inspecting Table 8 we see this is not the case, for example $\mathbf{X}_2$ mixes with $\mathbf{V}_2 + \mathbf{W}_2$. A likely explanation for mixing of torsional and spheroidal mode types lies in the boundary conditions. The boundary condition couples torsional and spheroidal modes whenever the surface normal $\hat{\mathbf{n}}$ is not in the radial direction $\hat{\mathbf{r}}$. Furthermore, as a vector $\hat{\mathbf{r}}$ carries angular momentum $l' = 1$. Hence the $\mathbf{X}_3$ mode can mix with mode $\mathbf{V}_4$, etc.

2.4 DISCUSSION

Having compared vibrational modes and eigenvalue spectra from simplified toy models and three different protein assemblies (LS, STMV and BMV), we note that the global modes in all cases may be expressed as combinations of vector spherical harmonics. These modes, which are likely to be biologically relevant, can be predicted and classified purely on the basis of molecular symmetry, which explains their distinctive shapes, and the strong resemblence between modes of

different structures. Given the values of two elastic moduli we can predict the vibrational frequencies of many low-lying modes. For spherical symmetry these predictions are highly accurate, and the accuracy remains high even for biological assemblies with icosahedral symmetry.

For example, the lowest frequency mode for all hollow structures turns out to be five-fold degenerate, with vector spherical harmonic type $\mathbf{V}_2 + \mathbf{W}_2$ (icosahedral type H). This mode can be visualized physically as the result of squeezing a sphere radially inwards at the poles, allowing it to bulge outwards at the equator. Such a deformation occurs when a molecule is probed with an atomic force microscope, and the hollow spheres are quite soft in response to this force [125, 126]. The applied uniaxial strain of the AFM tip couples to the mode of type H just discussed (superconductivity in the Fullerenes is also related to this mode [127, 128]).

To estimate the effective spring constant of this global deformation, we may express the energy as

$$E = \frac{M}{2}\lambda_H a_H^2 - a_H P_H F \tag{2.26}$$

where M represents the mass of the capsid, a_H is the amplitude of deformation in mode H and $P_H F$ is the projection of the applied force F onto the deformation $\mathbf{v}^{(H)}$. That is, $P_H F = \sum_i \mathbf{F}_i \cdot \mathbf{v}_i^{(H)}$ where $\mathbf{F}_i$ is the force acting on the i^{th} node and $\mathbf{v}^{(H)}$ is assumed normalized to 1. The first term in Eq. (2.26) represents deformation energy, while the second represents the work done by the applied force. Minimizing the energy yields the displacement in the direction of the applied force as $h = P_H a_H = P_H^2 F / M\lambda_H$ and an effective spring constant

$$k_{eff} = \frac{\partial F}{\partial h} = \frac{M\lambda_H}{P_H^2}. \tag{2.27}$$

This estimate is actually an upper bound on k_{eff} because other modes can share in the deformation energy if the applied force projects onto them also. The more general result is [120]

$$1/k_{eff} = \sum_k \frac{P_k^2}{M\lambda_k}, \tag{2.28}$$

where P_k represents the projection onto the eigenvector $\mathbf{v}^{(k)}$. Recently, Eyal and Bahar explored the single molecular response of external forces with elastic network models [129].

Next lowest in frequency (for our hollow sphere, LS and BMV) is the five-fold degenerate mode of vector spherical harmonic type $\mathbf{X}_2$ and icosahedral type H. This mode can be visualized as the result of twisting the upper and lower hemispheres in opposite directions.

For empty and filled STMV, modes of vector spherical harmonic type $\mathbf{V}_3 + \mathbf{W}_3$ (icosahedral types F_2 and G) appear at frequency below $\mathbf{X}_2$. These modes are poorly fit by the continuum elasticity theory and involve disordered protein chains (and nucleic acids) in the interior of the sphere.

Deformations of vector spherical harmonic type $\mathbf{V}_0$ (icosahedral type A) are nondegenerate because they preserve icosahedral symmetry. They correspond to radially directed displacements of angle-independent magnitude - a shrinking or swelling (breathing) of the entire structure.

For a solid sphere the relative frequencies of the lowest two mode types are interchanged relative to their order in the hollow sphere. As shown in Table 2 for the solid sphere, the eigenvalue of the spheroidal squeezing modes $\mathbf{V}_2 + \mathbf{W}_2$ (modes 6-10) become higher than the torsional twisting modes $\mathbf{X}_2$ (modes 1-5), reflecting the increased resistance to compression of the material near the center of the sphere. Similarly, in the solid sphere, the spheroidal $l = 1$

modes (Table 2, modes 11-13) appear at lower frequency than the $l = 3$ modes, while in the hollow sphere the $l = 1$ modes (Table 4, modes 18-20) appear above the frequency of the $l = 3$ modes. In the solid sphere these $l = 1$ modes correspond to an "optical" phonon in which the heavy interior of the sphere displaces in the opposite direction from the surface. In the hollow sphere (see Figure 1) these modes involve oppositely directed displacements of the poles and equator.

Although neither the ANM method, nor the continuum elasticity theory, can predict either the absolute value of the ANM force constant γ, or the continuum Young's modulus E, our fitting method yields ratios E/γ so that knowledge of one allows for the prediction of the other. For example, a γ = 0.1 N/m implies a Young's modulus of $E = 3.28$ GPa for the hollow sphere.

In the absence of experimental knowledge on elastic constants (or, equivalently, sound speeds), we could obtain the elastic constants by fitting to vibrational frequencies calculated on the basis of an intermolecular force field [130]. In this case, the continuum theory is no longer needed to *predict* vibrational frequencies, however it is still useful for the purpose of classifying the calculated normal modes. Additionally, the elastic constants obtained are needed for other applications of the continuum elasticity theory, such as finite element analysis of deformation.

Simple scaling laws follow from the continuum theory. For example, for a set of biomolecular assemblies sharing a common shape and elastic constants but differing in size, our theory predicts that the allowed wavenumbers q and k vary as the inverse of the linear dimension. Consequently, the lowest vibrational frequency also varies as the inverse of the linear dimension.

Our analysis so far deals with spherical and icosahedral structures. Similar ideas could be applied to other symmetries. For example, the modes in cylindrical molecules should be described as plane waves along the axis of the cylinders, sinusoidal oscillations around the circumference, and cylindrical Bessel functions in the radial direction. Such an analysis has been carried out for cylindrical viruses [131] allowing predictions of Raman vibrational frequencies.

In summary, our method provide a means of identifying and classifying the types of normal modes which arise from the structural symmetry, by expressing them in terms of icosahedral and vector spherical representations. The fitting procedure yields a relationship between the Young's modulus (E), which is a macroscopic elastic constant, and the microscopic force constant (γ) representative of inter-residue interactions. In addition, it yields the optimal Poisson's ratio (σ) that depends exclusively on the topological properties of the elastic network.

The present continuum model is restricted to the study of symmetric movements undergone by isotropic materials. Recent molecular simulations of STMV and BMV [132, 133] and nanoindentation experiments with viral capsids [134-136] exhibited asymmetric fluctuations and inhomogeneities in mechanical properties [137], which are beyond the scope of our theory.

3.0 ALLOSTERIC TRANSITIONS OF SUPRAMOLECULAR SYSTEMS EXPLORED BY NETWORK MODELS: APPLICATION TO CHAPERONIN GROEL[2]

Most proteins are biomolecular machines. They perform their function by undergoing changes between different structures. Understanding the mechanism of transition between these structures is of major importance to design methods for controlling such transitions, and thereby modulating protein function. While there are many computational methods for exploring the transitions of small proteins, the task of exploring the transition pathways becomes prohibitively expensive in the case of supramolecular systems. The bacterial chaperonin GroEL is such a supramolecular machine. It plays an important role in assisting protein folding. During its function, GroEL undergoes structural transitions between multiple forms. In this chapter, we are introducing a new methodology, based on elastic network models, for elucidating the transition mechanisms in such supramolecular systems. Application to GroEL provides us with biologically significant information on critical interactions and sequence of events occurring during the chaperonin machinery and key contacts that make and break at the transition. The method can be readily applied to other supramolecular machines.

[2] The content of this chapter has been published on *PLoS Comp. Bio.*, 2009 [103].

3.1 INTRODUCTION

3.1.1 Bacterial chaperonin GroEL

Many proteins assume more than one functional conformation, stabilized by ligand binding or changes in environmental conditions. A typical example is the bacterial chaperonin GroEL [138], a widely studied ATP-regulated molecular machine and member of heat shock protein Hsp60 family. GroEL assists in unfolding and refolding misfolded or partially folded proteins [139-141]. It is composed of two back-to-back stacked rings, each containing seven subunits of 60kDa. Each subunit is, in turn, composed of three domains, equatorial (E), intermediate (I) and apical (A). GroEL works together with the co-chaperonin, GroES.

The activity of GroEL-GroES complex entails a series of allosteric transitions in structure, triggered by ATP binding and hydrolysis, described in Figure 7: In the absence of nucleotide binding, both rings assume the closed (T) state, designated as T/T state for the two rings. Cooperative binding of seven ATP molecules to the subunits in one of the rings, hereafter referred to as the *cis* ring, drives the conformational change of these subunits to the 'open' (R) state, thus leading to the R/T form of the *cis/trans* rings. The R/T form exposes a number of hydrophobic residues at the domain A of the *cis* ring subunits. These groups attract the unfolded or partially folded peptide (substrate) to be encapsulated in the cylindrical chamber following the attachment of GroES (R'/T form of *cis/trans* rings). ATP hydrolysis provides the energy needed to process (unfold/refold) the substrate and leads to the state R"/T. This process is terminated upon binding of seven ATP molecules to the adjoining (*trans*) ring, hence the term 'negative cooperative effect' induced by ATP binding. The structure with ADP and ATP molecules bound to the respective *cis* and *trans* rings (R"/R form) favors the opening of the GroES cap and release

of the peptide and ADP molecules to start a new cycle, this time, with the roles of the former *cis* and *trans* rings being inverted.

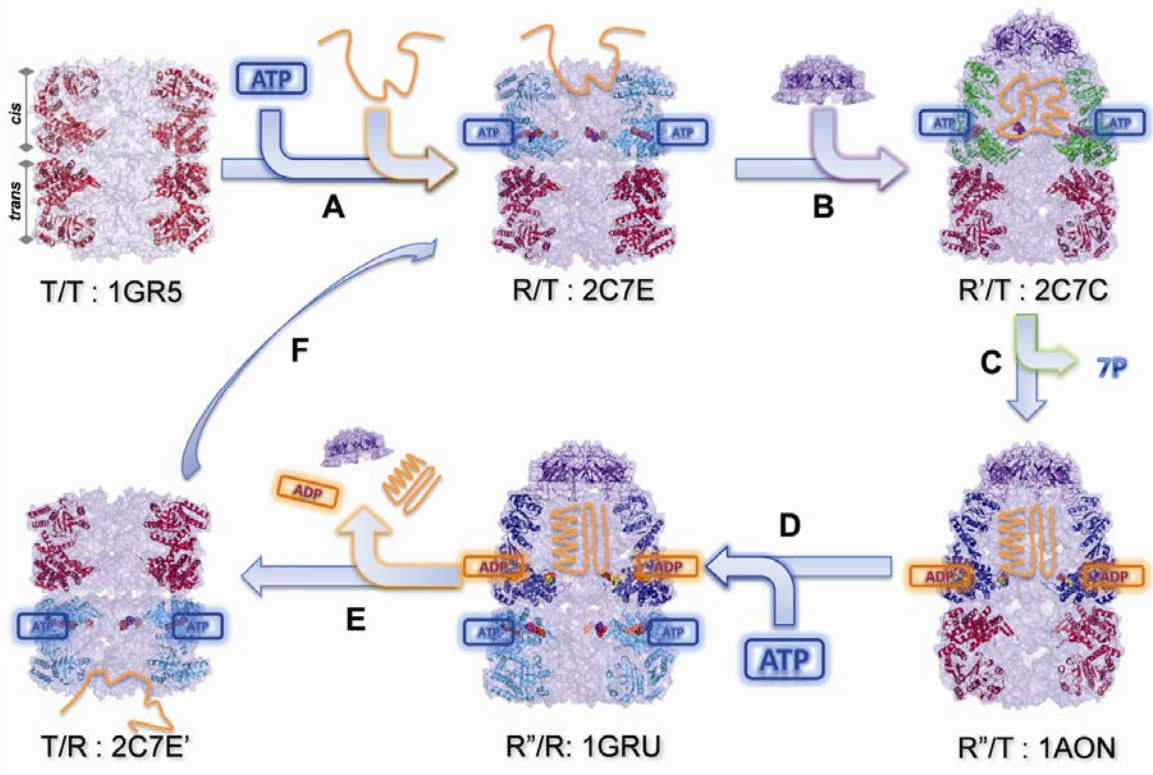

Figure 7. GroEL/GroES allosteric cycle.

GroEL consists of two rings, *cis* and *trans*, which assume the states: T: ATP-free; R: ATP-bound prior to substrate (peptide) and co-chaperonin (GroES) binding; R': ATP-, substrate- and GroES-bound; R": ADP-, substrate- and GroES-bound. Subunits in the T state are shown in red, R in cyan; R' in green, R" in blue, and the cap in purple. ATP and ADP are shown by blue and orange boxes. Successive events/reactions along the cycle are **(A)** binding of seven ATPs to induce the binding of the unfolded substrate (orange), **(B)** co-chaperonin binding, **(C)** ATP hydrolysis, **(D)** ATP binding to *trans* ring subunits, **(E)** release of ADPs, substrate (folded or partially folded) and GroES from the *cis* ring, **(F)** initiation of a new cycle where the roles of the *cis* and *trans* rings are inverted. Top-middle and bottom-left structures are related by rigid body rotation. Diagrams were generated using the data from the PDB in PyMOL (http://www.pymol.org), except for the schematic views of the substrate and ligands included to provide a clearer description. The PDB ids for the structures T/T, R/T, R"/R, R'/T and R"/T are 1GR5, 2C7E and 1GRU [142], 2C7C [143] and 1AON [144], respectively.

Of interest is to understand the molecular basis of the negative cooperative effect triggered upon binding of seven ATP molecules to the *trans* ring (step **E** in Figure 7). ATP binding induces in this case a structural change on the cap-binding region at a distance of 65 Å. Understanding the mechanism of this allosteric signaling is of fundamental importance because of the critical role chaperonins play in preventing aggregation and regulating folding *vs.* degradation events.

Several studies have been published on the allosteric pathways and dynamics of GroEL [27, 32, 33, 140, 141, 143, 145-155] since the original determination of the ADP-bound complex (R"/T) [144]. Perhaps the most comprehensive computational study of GroEL allosteric dynamics is that of Hyeon, Lorimer and Thirumalai, where the T → R → R" transition has been examined by Brownian dynamics (BD) simulations using a state-dependent (Go-like) self-organized polymer model [154]. These simulations, performed for subunit dimers and heptamers, provided valuable insights on the heterogeneity of the transition pathways and the kinetics of salt bridges' formation/rupture at the successive transitions T → R and R → R".

These studies provide valuable insights into the successive states involved in the chaperonin cycle. The elucidation of allosteric transition mechanisms has been a challenge, however, both experimentally and computationally, due to the transient nature of the high energy conformers between the states, the multiplicity of pathways, and the large size of this biomolecular system.

The following GroEL structures deposited in the Protein Data Bank (PDB) [3] have been considered in the present study (Figure 7): (**A**) 1GR5; the apo-form of GroEL, representative of the T/T state. (**B**) 1GRU; identified with the R"/R state; it contains seven ADPs and seven ATPs bound to the *cis* and *trans* rings, respectively. All these structures were determined by cryo-

electron microscopy by Saibil and coworkers [142, 143, 156]. The structures being compared (the end points and intermediates at each step) were optimally superimposed at their α-carbons using Kabsch method [157] prior to implementing the *a*ANM method.

3.1.2 Assumptions

The ANM is based on the following assumptions: (i) the experimentally resolved structures are assumed to represent an energy minimum, (ii) the structures are modeled as a network, the nodes of which represent the individual amino acids, (iii) the positions of the nodes are identified by those of the α-carbons, (iv) all pairs of nodes separated by a distance shorter than a cutoff distance of r_c are connected by an elastic spring, (v) a uniform force constant is assumed for all pairs regardless of amino acid type. Thus, the structure is subject to a sum of harmonic potentials corresponding to all connected (bonded or non-bonded) amino acids. The structural coordinates are used to derive the *3N-6* orthonormal modes of motion, called ANM modes, uniquely defined for each equilibrium structure of *N* modes [20, 21]. The ANM is now widely used in exploring protein dynamics and assessing the global (lowest frequency modes) in particular, in view of the observed robustness and functional relevance of these modes [14-18, 42, 43, 158] . More details on the model and method may be found in section 1.2.2.

In the present study, the potential of the system is approximated by the sum of two harmonic ANM potentials adopted for the end states, along with a softening term that ensures a smooth transition between them. As will be explained below, the softening term (and the corresponding parameter β; see Methods) is not used in the *a*ANM generation of the pathways, but in the comparison of the results against those predicted by other methods.

Multiple paths (or subsets of ANM modes) exist for the passage between the endpoints. The recruitment of the particular subsets of modes results from a tradeoff between minimizing the path length and selecting the direction of the lowest increase in internal energy. The 'shortest' path is by definition the interpolation between the two endpoints. However, this path usually incurs unphysical strains in internal coordinates, and is, thereby, unfavorable from energetic point of view. In the other extreme case of movements along the lowest-lying (softest) modes, on the other hand, these particular mode directions may not necessarily lead to the target. The selected pathway will thus depend on the relative importance of environmental (*path length*) or internal (*conformational energy barrier*) constraints (see below).

The results are organized in two sections: First (*Section 3.3.1*), we present the results for the R" ←→ T transition of a single GroEL subunit (N = 514 residues). These calculations serve two purposes: benchmarking our results against those obtained by action minimization method [59], and estimating the effect of different choices of parameters. These data are then utilized in the second part (*Section 3.3.2*) where we examine the intact chaperonin allosteric cycle and compare with experimental data and other simulations. Notably, a small subset of low frequency modes are found to drive the transition at early stages, succeeded by the recruitment of increasingly larger subsets of modes (yet not exceeding 4% of the entire spectrum) to overcome the energy barrier. Native contacts are closely maintained throughout significant portions of transition pathways, except for the close neighborhood of the energy peak, also called putative transition state (PTS), where a major redistribution in intra- and inter-subunit contacts takes place. The results provide new insights and testable hypotheses on the mechanistic involvement of conserved residue pairs in critical interactions.

3.2 THEORY AND METHOD

3.2.1 Definitions

The method is based on the simultaneous generation of pairs of intermediate conformations, $\boldsymbol{R}_A^{(k)}$ and $\boldsymbol{R}_B^{(k)}$, starting from the known endpoints $\boldsymbol{R}_A^{(0)}$ and $\boldsymbol{R}_B^{(0)}$. Each conformation along the so-called reaction coordinate is represented by a *3N*-dimensional vector, corresponding to the coordinates of the α-carbons. The *distance vector* $\boldsymbol{d}^{(k)}$ between the pair of conformations generated at iteration *k* is defined as

$$\boldsymbol{d}^{(k)} = \boldsymbol{R}_B^{(k)} - \boldsymbol{R}_A^{(k)}, \tag{3.1}$$

and the *deformation vectors* used to generate the k^{th} pair of conformations are

$$\begin{aligned} \boldsymbol{v}_A^{(k)} &= \boldsymbol{R}_A^{(k)} - \boldsymbol{R}_A^{(k-1)} \\ \boldsymbol{v}_B^{(k)} &= \boldsymbol{R}_B^{(k)} - \boldsymbol{R}_B^{(k-1)}. \end{aligned} \tag{3.2}$$

The original distance vector between the two optimally superimposed end structures is $\boldsymbol{d}^{(0)} = \boldsymbol{R}_B^{(0)} - \boldsymbol{R}_A^{(0)}$. This distance vector also defines the original root-mean-square deviation (RMSD), $|\boldsymbol{d}^{(0)}|/N$, between the endpoints. The RMSD at iteration *k* is similarly defined as

$$\mathrm{RMSD}(\boldsymbol{R}_A^{(k)}, \boldsymbol{R}_B^{(k)}) = \left|\boldsymbol{d}^{(k)}\right| / N \tag{3.3}$$

3.2.2 Methodology

The ***a*ANM** method consists of the following steps (see Figure 8):

(i) Two sets of intermediates are generated, starting from both ends. The recurrence equation for evaluating the k^{th} intermediate starting from state *A* is

$$\boldsymbol{R}_A^{(k)} = \boldsymbol{R}_A^{(k-1)} + \boldsymbol{v}_A^{(k)} = \boldsymbol{R}_A^{(k-1)} + s_A^{(k)} \sum_{i=1}^{m_A^{(k)}} \left(\boldsymbol{d}^{(k-1)} \cdot \boldsymbol{u}_{iA}^{(k)} \right) \boldsymbol{u}_{iA}^{(k)} \tag{3.4}$$

and a similar expression holds for state *B*. Following Eq. (3.4) , we evaluate $\boldsymbol{v}_A^{(k)}$ on the basis of $m_A^{(k)}$ dominant (low frequency) eigenvectors $\boldsymbol{u}_{iA}^{(k)}$ ($1 \le i \le m_A^{(k)}$) predicted by the ANM for the conformation $\boldsymbol{R}_A^{(k-1)}$. The displacement along the i^{th} eigenmode is proportional to the projection $\boldsymbol{d}^{(k-1)} \cdot \boldsymbol{u}_{iA}^{(k)}$ of the instantaneous distance vector $\boldsymbol{d}^{(k-1)}$ on $\boldsymbol{u}_{iA}^{(k)}$, and scaled by the *step size* $s_A^{(k)}$. (The step sign has been included in the projection $\boldsymbol{d}^{(k-1)} \cdot \boldsymbol{u}_{iA}^{(k)}$.) The step sizes $s_A{}^{(k)}$ and $s_B{}^{(k)}$ are simultaneously selected at each iteration *k*, as a fraction *f* of those, $s_{A,m}{}^{(k)}$ and $s_{Bm}{}^{(k)}$, that minimize $\boldsymbol{d}^{(k)}$. The limit $f \rightarrow 0$ refers to infinitesimally small displacements that are strictly accurate but prohibitively expensive, while the other extreme case $f \rightarrow 1$ is the most efficient move, but may give rise to unphysical deformations in structural coordinates. f = 0.2 is selected here as a scaling factor that optimally balances between efficiency and accuracy (see *Methods*).

(ii) The number $m_A^{(k)}$ of modes of motion recruited at iteration *k* is based on a threshold squared cosine, F_{min}, that defines the maximal angular departure between the instantaneous displacement direction and that targeted. To this end, we evaluate the *cumulative squared cosine*

$$\left[C\left(m_A^{(k)} \right) \right]^2 = \sum_{i=1}^{m_A^{(k)}} cos^2 \left(\boldsymbol{d}^{(k-1)}, \boldsymbol{u}_{iA}^{(k)} \right), \tag{3.5}$$

and we select the minimal number of modes, starting from the low frequency end of the spectrum, that satisfy the inequality $[C(m_A{}^{(k)})]^2 \ge F_{min}$. It can be shown (see derivation in Appendix C) that $C(m_A{}^{(k)})$ is identical to the correlation cosine between the instantaneous deformation and distance vectors, i.e.,

$$C\left(m_A^{(k)} \right) = cos(\boldsymbol{d}^{(k-1)}, \boldsymbol{v}_A^{(k)}). \tag{3.6}$$

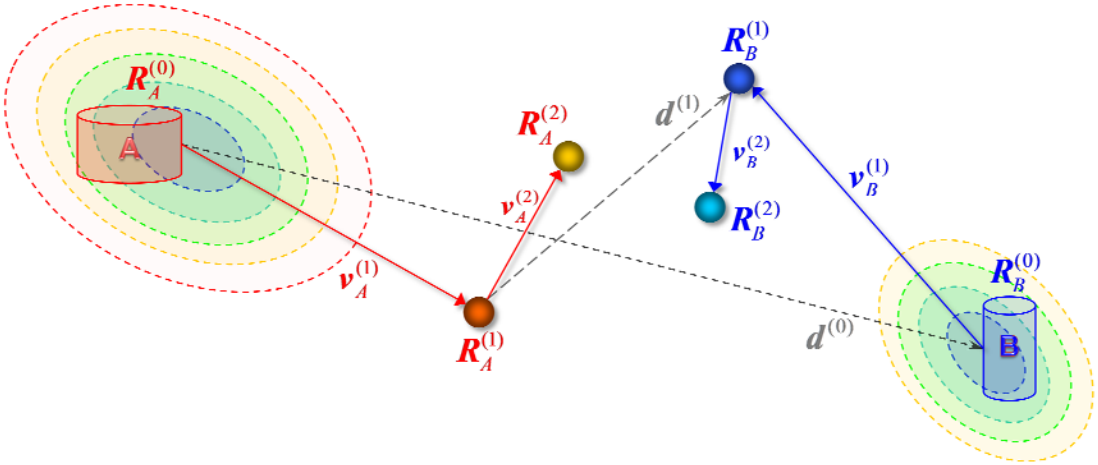

Figure 8. Schematic description of *a*ANM method.

Two sets of intermediate conformations are generated, $\boldsymbol{R}_A^{(k)}$ and $\boldsymbol{R}_B^{(k)}$ ($1 \le k \le k_{tot}$), starting from the known substates $\boldsymbol{R}_A^{(0)}$ and $\boldsymbol{R}_B^{(0)}$, illustrated here for $k = 1$ and 2. The distance vector between the instantaneous endpoints at the k^{th} step is denoted as $\boldsymbol{d}^{(k)}$, and the deformation at each step is $\boldsymbol{v}_A^{(k)}$ or $\boldsymbol{v}_B^{(k)}$. Dashed ellipses indicate isoenergetic contours.

Therefore, the threshold F_{min} ensures the selection of the *smallest* subset of modes to drive the deformation $\boldsymbol{v}_A^{(k)}$ of the molecule toward a direction that does not deviate by more than a specified correlation cosine ($F_{min}^{½}$) from the target direction $\boldsymbol{d}^{(k-1)}$. Note that the use of the complete set of modes leads, by definition, to $C(3N\text{-}6) = 1$. By selecting a subset, we let the molecule undergo a structural change that is not necessarily toward the endpoint, but along the coordinates energetically favored by its architecture.

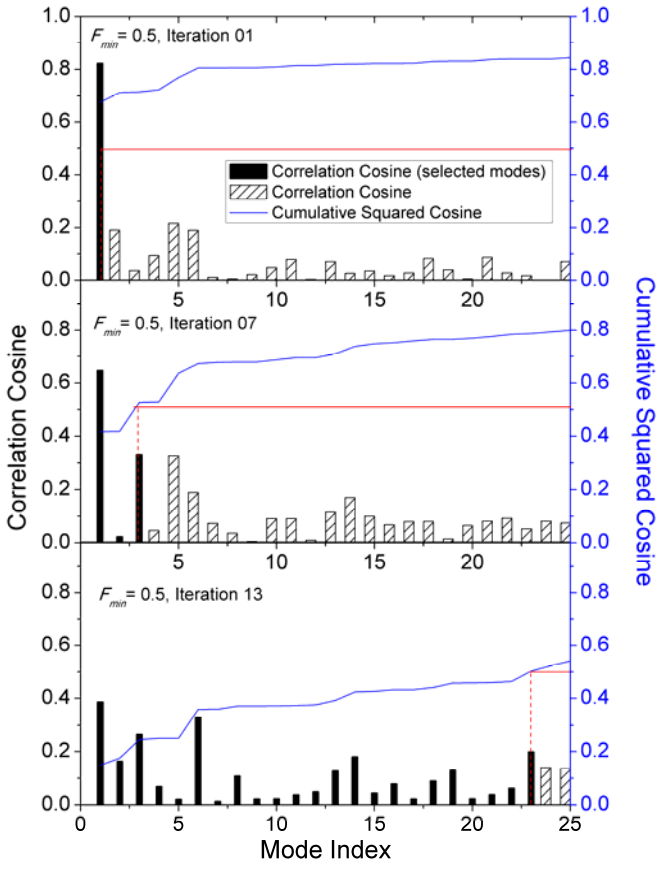

Figure 9. Correlation cosine between instantaneous distance vector and eigenmodes.

Results are illustrated for *a*ANM steps k = 1, 7 and 13 along the transition R" → T of a single subunit (subunit A in the respective PDB structures 1GRU and 1GR5). The left ordinate displays the correlation cosine between the distance vector $\boldsymbol{d}^{(k-1)}$ and the eigenvectors $\boldsymbol{u}_{iA}^{(k)}$ for $1 \le i \le 30$ (black bars), and the right ordinate shows the corresponding cumulative squared cosine (Eq. (3.5)) (blue curve). The threshold F_{min} = 0.5 for the cumulative square cosine implies the selection of $m_A^{(1)} = 1$, $m_A^{(7)} = 3$, and $m_A^{(13)} = 23$ in evaluating $\boldsymbol{v}_A^{(k)}$ as indicated by the red lines and filled bars. See Table 9 for the complete list of $m_A^{(k)}$ and $m_B^{(k)}$ values and associated RMSDs between intermediate conformations.

Figure 9 illustrates the procedure for selecting $m_A^{(k)}$ for the transition R" → T (step **E** in Figure 7). Results are shown for the *a*ANM iterations k = 1 (*top*), 7 (*middle*) and 13 (*bottom*). The bars displays the correlation cosine $cos(\boldsymbol{d}^{(k-1)}, \boldsymbol{u}_{iA}^{(k)})$ as a function of mode number $1 \le i \le 25$

(*left* ordinate), and the blue curve is the cumulative squared cosine $[C(m_A^{(k)})]^2$ (*right*). For $k = 1$, the lowest frequency mode ($i = 1$) alone yields a correlation cosine of 0.82: it suffices, therefore, to take $m_A^{(1)} = 1$ mode at this step to meet the criterion $[C(m_A^{(k)})]^2 \geq F_{min}$, if the threshold $F_{min} = 0.5$. For $k = 7$, on the other hand, the same criterion is met by $m_A^{(7)} = 3$ modes (see the red line), and for $k = 13$, we need $m_A^{(13)} = 23$ modes.

Table 9 summarizes the $m_A^{(k)}$ and $m_B^{(k)}$ values for all steps along the R" → T transition for $F_{min} = 0.4$, 0.5, 0.6 and 0.7. Interestingly, an increasingly larger number of modes are recruited as we move away from the original equilibrium state. This important result, also confirmed for the intact complex, will be further elaborated below.

The ***a*ANM** calculations thus involve two parameters, F_{min} and f. The former controls the direction of motion, and the latter its size. Smaller F_{min} values permit us to proceed via lower energy ascent directions, at the cost of longer excursions. Smaller f implies smaller displacements at each iteration.

(iii) The above scheme is repeated to generate a series of intermediate conformations, until the RMSD between the intermediates becomes sufficiently small (1.5 Å in Table 9). The total number of iterations, k_{tot}, is thus defined by this targeted RMSD. The sequence of conformations along the pathway is denoted as $\left\{\boldsymbol{R}_A^{(0)}, \boldsymbol{R}_A^{(1)} \cdots \boldsymbol{R}_A^{(k_{tot})}, \boldsymbol{R}_B^{(k_{tot})} \cdots, \boldsymbol{R}_B^{(1)}, \boldsymbol{R}_B^{(0)}\right\}$ or $\left\{\boldsymbol{R}^{(1)}, \boldsymbol{R}^{(2)}, \cdots, \boldsymbol{R}^{(n)}\right\}$.

Table 9. *a*ANM data for the transition of a GroEL subunit between R" and T forms(∗)

Iteration (k)	F_{min} = 0.4			F_{min} = 0.5			F_{min} = 0.6			F_{min} = 0.7		
	$m_{R"}^{(k)}$	$m_T^{(k)}$	*RMSD* (Å)	$m_{R"}^{(k)}$	$m_T^{(k)}$	*RMSD* (Å)	$m_{R"}^{(k)}$	$m_T^{(k)}$	*RMSD* (Å)	$m_{R"}^{(k)}$	$m_T^{(k)}$	*RMSD* (Å)
0			12.33			12.33			12.33			12.33
1	1	3	11.00	1	4	10.74	1	4	10.74	2	6	10.56
2	1	3	9.77	1	3	9.57	1	4	9.46	2	4	9.16
3	1	3	8.80	1	3	8.48	1	3	8.37	2	4	7.97
4	1	3	8.01	1	3	7.61	1	4	7.37	3	4	6.96
5	1	3	7.36	1	3	6.93	3	5	6.42	5	6	6.12
6	1	3	6.77	3	4	6.16	3	4	5.70	5	7	5.29
7	3	4	6.04	3	4	5.57	5	4	5.11	6	7	4.66
8	3	4	5.46	5	4	5.09	6	4	4.53	13	7	4.14
9	3	4	5.01	6	6	4.50	11	5	4.04	23	12	3.68
10	5	4	4.58	6	7	3.98	15	6	3.60	41	16	3.27
11	6	6	4.18	13	7	3.56	23	10	3.26	73	33	2.91
12	6	7	3.73	17	7	3.21	32	16	2.96	112	44	2.59
13	10	7	3.34	23	7	2.96	64	27	2.68	210	119	2.31
14	14	7	3.06	31	16	2.70	87	39	2.43	377	259	2.06
15	17	7	2.82	54	22	2.49	156	72	2.20	576	455	1.83
16	24	12	2.62	75	35	2.30	266	169	1.99	906	727	1.63
17	32	17	2.44	111	43	2.12	422	312	1.81	1189	1118	0.87
18	52	25	2.27	164	89	1.95	651	488	1.64	-	-	-
19	72	35	2.11	267	172	1.80	953	742	1.49	-	-	-
20	98	41	1.98	385	295	1.66	-	-	-	-	-	-
21	137	68	1.86	545	442	1.53	-	-	-	-	-	-
22	192	124	1.74	769	639	1.05	-	-	-	-	-	-
23	284	207	1.64	-	-	-	-	-	-	-	-	-
24	388	304	1.54	-	-	-	-	-	-	-	-	-
25	498	443	1.44	-	-	-	-	-	-	-	-	-

(∗) $m_T^{(k)}$ and $m_{R"}^{(k)}$ are the number of modes recruited at step k, starting from the respective states *R"* and *T; RMSD* values are calculated using Eq. (3.3). A and B designate the respective states T and R". The first row indicates the original RMSD of 12.33 Å between the two end points.

3.2.3 *a*ANM method and parameters

To explore the transition pathway between the initial and final conformations of subunit A, a series of intermediate conformations have been generated upon successive deformations of both end structures that were regularly updated. The directions of deformations were determined by implementing the deformations along the directions of dominant ANM modes accessible to the intermediate states. The contribution of the i^{th} normal mode to the mean-square fluctuations of residues scales with the inverse eigenvalue of the Hessian $\boldsymbol{H}$ characteristic of the energy landscape near the equilibrium state; and the eigenvalues, in turn, scale with the squared frequency of the normal modes, hence the dominant role of low frequency modes in shaping the equilibrium dynamics [15].

The *a*ANM scheme involves two parameters, F_{min} and f, which, in turn, determine $m_A^{(k)}$, $m_B^{(k)}$, $s_A^{(k)}$ and $s_B^{(k)}$. The dependence of $m_A^{(k)}$ and $m_B^{(k)}$ on F_{min} has been described in the Theory. As to $s_A^{(k)}$ and $s_B^{(k)}$, these are determined as follows: first we evaluate the values $s_{Am}^{(k)}$ and $s_{Bm}^{(k)}$ that minimize $|d^{(k)}|$. Note that there is a unique combination of these parameters (or the sizes $|v_A^{(k)}|$ and $|v_B^{(k)}|$ instantaneous displacements) which leads to the minimal $|d^{(k)}|$. However, the step sizes that meet this criterion may be too large to be extrapolated using a quadratic approximation away from the energy minimum, and the energy cost of the induced deformations may be excessively high and not accurately accounted for by the harmonic approximation. Instead, we take a fraction f of the calculated values, i.e., we implement displacements that scale with $fs_{Am}^{(k)}$ and $fs_{Bm}^{(k)}$. Calculations repeated with different f values support the use of relatively small values for two reasons: First, the smaller steps avoid unphysical distortions in the structure; second, if a larger displacement along a given mode direction is more productive, the same mode is selected in the succeeding iteration (see for example how modes 1, 3, 4 etc. are selected

multiple times at the initial steps of the trajectories in Table 9.) Therefore, in principle, we would like to take steps that are as small as possible. However, the simulations become increasingly time-consuming (due to larger number of iterations) with decreasing step size. Our calculations suggest that the value f = 0.2 leads to sufficiently small step sizes along with computationally tractable iterations.

Finally, instead of picking a fixed F_{min} throughout the entire trajectories, F_{min} can be chosen dynamically, $F_{min}^{(k)} = 1 - \sqrt{\left|\boldsymbol{d}^{(k-1)}\right| / \left|\boldsymbol{d}^{(0)}\right|}$ (for $k > 1$). In the first iteration, only the first dominant mode is used to deform the system. The results with this choice of dynamic $F_{min}^{(k)}$ were found to yield results comparable to F_{min} = 0.4- 0.5 and SDP results. The use of this empirical relationship has the advantage of eliminating the variable F_{min} from the *a*ANM algorithm.

3.2.4 Potential energy function and parameters

We adopt a double minima quadratic energy function [94]

$$U = \frac{1}{2}\left(U_A + U_B - \sqrt{\left(U_A - U_B\right)^2 + 4\beta^2} \right). \tag{3.7}$$

Here the parameter β sets the height (and smoothness) of the energy barrier, and the potentials U_A and U_B are defined in the ANM in terms of a uniform force constant γ as [21]

$$\begin{aligned} U_A &= \frac{\gamma}{2} \sum_{r_{ij} < r_c} \left(\left|\boldsymbol{r}_{ij}\right| - \left|\boldsymbol{r}_{ij}^{A}\right| \right)^2 \\ U_B &= \frac{\gamma}{2} \sum_{r_{ij} < r_c} \left(\left|\boldsymbol{r}_{ij}\right| - \left|\boldsymbol{r}_{ij}^{B}\right| \right)^2 . \end{aligned} \tag{3.8}$$

Here $|\boldsymbol{r}_{ij}|$ is the instantaneous distance between residues i and j (based on α-carbons), $|\boldsymbol{r}_{ij}^{A}|$ is the equilibrium distance in conformation A, and the summations are performed over all pairs of residues located within a cutoff distance r_c. The *a*ANM trajectories were generated using Eq. (3.8) in each iteration by assuming that $\boldsymbol{r}_{ij}^{A}$ and $\boldsymbol{r}_{ij}^{B}$ are the final conformations of the previous iteration. Eq. (3.7) has been used for evaluating the energy of the generated conformations and for performing SDP calculations. We adopted the value $\beta = 10$ kcal/mol and verified that changes in β within one order of magnitude do not affect the results. Nevertheless, the energies reported below provide information on the shape, rather than absolute values, of the energy profile. The force constant is taken as 0.7 kcal/(mol·Å^2), and the cut-off distance as r_c = 13.0 Å consistent with previous assessment of optimal ANM parameters [158].

3.3 RESULTS

3.3.1 Allosteric transitions of a single subunit

The allosteric cycle undergone by a given subunit can be summarized as T → R → R' → R" → T. Table 10 **A** lists the RMSDs between these alternative states, derived from the coordinates of the *cis* ring subunits in different forms of the complex. The RMSD between R' and R" is 1.50 Å, which is lower than the resolution of these structures, while those between different states are larger than 12 Å, except for the T and R states which differ by 5.21 Å. A reasonable approximation in view of the similarity of R' and R", and the resolution of existing structural data, is to condense the allosteric cycle into T → R → R" (or R') → T (see for example [154]).

Toward exploring the suitability of *a*ANM for predicting the events T → R → R" → T, we first examine the *intrinsic ability* of the subunits in the T, R and R" forms to undertake these particular changes (steps **B**, **D** and **E**, in Figure 7). Second, we focus on the transition R" → T, before proceeding to the intact chaperonin in section II.

Table 10. Structural differences (RMSD values in Å) between (A) different forms of a subunit and (B) different states of the intact GroEL.

A. Single GroEL subunit (*)					
RMSD (Å)	1GR5_A (**T**)	2C7E_A (**R**)	2C7C_A (**R'**)	1GRU_A (**R"**)	
1GR5_A (**T**)	--	5.21	12.06	12.33	
2C7E_A (**R**)	5.21	--	12.23	12.35	
2C7C_A (**R'**)	12.06	12.23	--	1.50	
1GRU_A (**R"**)	12.33	12.35	1.50	--	
B. Intact GroEL					
RMSD (Å)	1GR5 (**T/T**)	2C7E (**R/T**)	2C7C (**R'/T**)	1AON (**R"/T**)	1GRU (**R"/R**)
1GR5 (**T/T**)	--	6.72	12.09	11.48	12.19
2C7E (**R/T**)	6.72	--	10.85	10.65	11.03
2C7C (**R'T**)	12.09	10.85	--	3.17	2.65
1AON (**R"/T**)	11.48	10.65	3.17	--	2.67
1GRU (**R"/R**)	12.19	11.03	2.65	2.67	--

(*) based on α-carbons. The states refer to *cis* ring subunit A (indicated by suffix _A).

3.3.1.1 Intrinsic dynamics of T, R and R" states favor functional changes in structure

The lowest frequency modes, also called *global modes*, represent by definition the collective motions that are most easily accessed near a given equilibrium state, or those intrinsically favored by the 3D structure [18]. These are soft modes along entropically preferred reconfigurations [11]. We will analyze each of the three steps in T → R → R" (or R') → T to see to what extent the corresponding changes in structure, which are functionally required, comply with these modes.

Results are summarized in Table 11. The structural changes induced by the slow modes allow for substantial decrease in the distance (RMSD) from the endpoint except for mode 1 for the step R→R" (*middle* panel on the *left*) where the mode direction is apparently almost orthogonal to $\boldsymbol{d}^{(0)}$. Table 11 lists the closest RMSDs that can be achieved upon moving along one, two, or three low frequency modes, starting from either end, or proceeding simultaneously from both ends. Strikingly, in all cases, the three modes permit the molecule to complete more than half of the structural change in the 'functional' direction, confirming that the subunits are intrinsically poised to reconfigure in the 'right' way. Below are more details. For clarity, the entries discussed below are highlighted (in boldface).

(i) *Step T → R.* The T state exhibits a strong tendency to approach the R state. The change is indeed a reconfiguration along the first mode predicted by the ANM for the T state. This type of transition occurs both in steps **A** and **D** of the allosteric cycle (Figure 7) at the respective *cis* and *trans* ring subunits. As the T subunit changes its conformation along mode 1 alone, the RMSD from state R decreases from 5.21 to 3.40 Å, and upon further recruiting modes 2 and 3, the RMSDs decrease to 3.05 and 2.83 Å, respectively. Interestingly, the end state (R) also tends to move toward the state T, if deformed along its mode 1. These results support the view that the T form possesses an intrinsic tendency to assume its ATP-bound conformation R, prior to ATP binding, and it is also readily recovered upon nucleotide release.

(ii) *Step R → R"* (steps **B** and **C** in Figure 7). In this case, the RMSD of 12.35 Å between the end states cannot be reduced upon moving along mode 1. Instead mode 2 appears to play a major role, to decrease the RMSD to 8.19 Å. This step is triggered upon ATP hydrolysis, which apparently favors mode 2. Yet, it is remarkable that the contribution of three modes from both ends is sufficient to reduce the RMSD to 4.19 Å.

Table 11. Contributions of the lowest frequency modes to the allosteric cycle T → R → R" → T of GroEL[(a)]

Step	Starting structure	Target structure	Original RMSD (Å)	RMSD by *the 1st* ANM mode			RMSD by *two* ANM modes[(b)]			RMSD by *three* ANM modes[(b)]		
				RMSD (Å)	F_{min}	modes	*RMSD* (Å)	F_{min}	modes	*RMSD* (Å)	F_{min}	modes
T to R	1GR5_A (**T**)	2C7E_A (**R**)	**5.21**	**3.40**	0.57	1	**3.05**	0.65	1+2	**2.83**	0.70	1+2+3
	2C7E_A (**R**)	1GR5_A (**T**)		**2.83**	0.70	1	2.70	0.73	1+4	2.65	0.75	1+4+6
	Deform simultaneously			2.74			2.61			**2.53**		
R to R"	2C7E_A (**R**)	1GRU_A (**R"**)	**12.35**	12.11	0.04	1	**8.19**	0.55	1+2	7.30	0.65	1+2+3
	1GRU_A (**R"**)	2C7E_A (**R**)		8.84	0.49	1	7.65	0.61	1+2	7.29	0.65	1+2+3
	Deform simultaneously			7.57			6.02			**4.19**		
R" to T	1GRU_A (**R"**)	1GR5_A (**T**)	**12.33**	**7.02**	0.68	1	6.61	0.71	1+2	6.05	0.76	1+2+5
	1GR5_A (**T**)	1GRU_A (**R"**)		12.03	0.05	1	8.78	0.49	1+3	7.38	0.64	1+3+4
	Deform simultaneously			6.49			6.19			**4.91**		

[(a)] The *RMSD*s refer to those between the intermediates generated by moving along the 1st ANM mode (columns 5-7), two modes (8-10) and three modes (11-13) in iteration $k = 1$ (see Eq. (3.3)). For each step (1st column), the results are separately given for the reconfiguration of the forward (first row), backward (2nd row) and simultaneous (3rd row) passages between the two endpoints.

[(b)] Lowest frequency modes that exhibit a correlation cosine of > 0.1 with $\boldsymbol{d}^{(0)}$ are selected. Note that these are all confined to the lowest frequency six ANM modes (see listed mode numbers).

(iii) *Step R"→ T* (step **E** in Figure 7). Results illustrated in Table 9 already showed that the reconfiguration of R" is initiated via a deformation along the first mode. This mode indeed yields a correlation cosine of 0.82 with the targeted direction $\boldsymbol{d}^{(0)}$ (see Figure 9), and Table 11 shows that the original RMSD of 12.33 Å decreases to 7.02 Å upon moving along mode 1, exclusively. As to the backward transition, the first mode of the T state induces a movement almost orthogonal to $\boldsymbol{d}^{(0)}$. Yet, recruitment of the first three modes from both ends has a dramatic effect, as an RMSD of 4.91 Å is reached.

The above analysis shows the utility of taking steps along the ANM modes so as to move efficiently toward the endpoint. Slowest modes refer here to lowest-lying eigenvalues, or softest modes. Recruiting three modes from both ends leads to intermediate states that differ by 2.53, 4.19 and 4.91 Å, as opposed to the original RMSDs of 5.21, 12.35 and 12.33 Å between the end points of the respective three steps T → R → R" → T. The extent of productive reconfiguration, or the fraction of the total path travelled (more than ½) by taking steps along these three modes is remarkable, given that (i) these represent 3/1,542 of the complete set of *3N-6* normal axes/directions that could be selected by the starting conformation, and (ii) the ANM modes are solely based on the distribution of inter-residue contacts.

3.3.1.2 R" → T transition analyzed by *a*ANM: path lengths *vs.* energy barriers

Having established the use of moving along ANM modes, we now proceed to an iterative use and (re)generation of ANM modes according to the *a*ANM algorithm. In order to assess the effects of the choice of *a*ANM parameters and establish default values, which will also be adopted for the intact GroEL (see below), we repeated *a*ANM calculations for different F_{min} values. Results for the transition R"→ T are shown in Figure 10**.** Panel **A** displays the gradual decrease in the RMSD between the instantaneous pairs of intermediate conformations (see Eq. (3.3)) as a function of iteration number *k,* and panel **B** the change in energy involved in each case. It can be seen that lower F_{min} values allow for larger excursions away from the targeted direction by recruiting relatively smaller numbers of low frequency modes. They consequently require a larger number of steps to be undertaken to reach the target, while the accompanying energy increase is relatively smaller. Higher F_{min}, on the other hand, permits us to reach the target faster, but with a higher energy cost. The limit $F_{min} = 1$ corresponds to pure interpolation by recruiting all modes.

The *reaction coordinate* in panel **B** is the projection of the cumulative displacement $\boldsymbol{v}^{(n)} = \boldsymbol{R}^{(n)} - \boldsymbol{R}_A^{(0)}$ on the original distance vector $\boldsymbol{d}^{(0)}$, i.e., $x(n) = \boldsymbol{d}^{(0)}.\boldsymbol{v}^{(n)}/|\boldsymbol{d}^{(0)}|^2$, with R" and T representing the respective limits $x(n) = 0$ and 1. The peak in the energy profile tends to be closer to the T state, especially when lower F_{min} values (which entail lower energy barriers) are adopted. This may be related to the recruitment of higher modes (steeper ascent along the energy surface) near state T, as opposed to the first mode near R". Note that in the allosteric transition of hemoglobin between the T and R' states, T was found to be more predisposed to move toward R", than R" undergoing the reverse transition [11, 159]. In particular, the value $F_{min} = 0.5$ leads to a peak around $x(n) = 0.52$, in accord with the SDP derived by action minimization (black curve in Figure 10**B**), as will be further elaborated below.

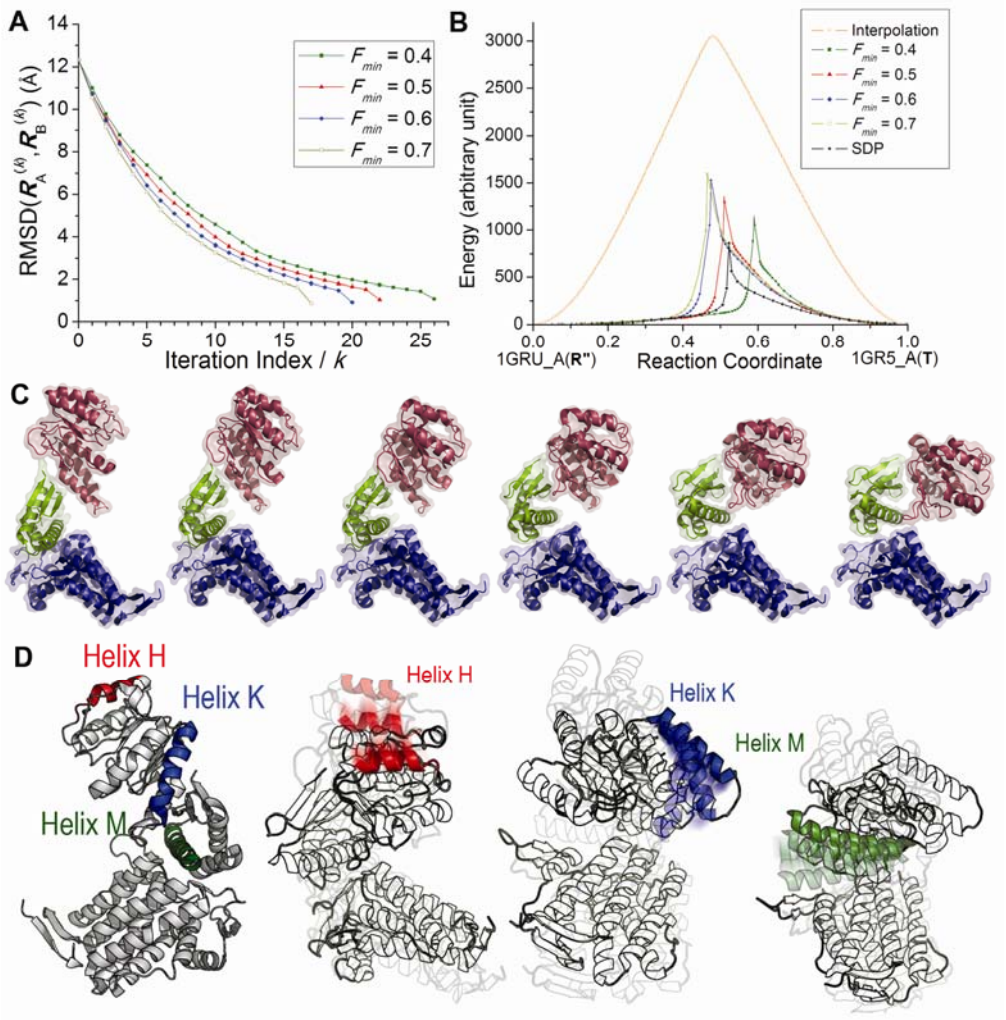

Figure 10. R" → T transition for a single subunit of GroEL.

(A) RMSD values, $|\boldsymbol{d}^{(k)}|$ /N, between instantaneous endpoints plotted as a function of iteration number k. The end states refer to subunit A in the PDB structures 1GRU and 1GR5. Results are shown for F_{min} = 0.4, 0.5, 0.6 and 0.7, corresponding to the allowable angular deviations of up to 50.8°, 45.0°, 39.2° and 33.2°, respectively, between $\boldsymbol{d}^{(k)}$ and $\boldsymbol{v}_A^{(k)}$ (or $\boldsymbol{v}_B^{(k)}$). **(B)** The energy profile for alternative pathways in arbitrary units. Note the significantly lower energy barrier compared to the interpolation (orange curve) between the endpoints. The black curve refers to the SDP trajectory. The reaction coordinate refers to the normalized projection of the instantaneous displacement on the original distance vector. **(C)** Series of conformations sampled along the reaction coordinate. The diagrams are colored by domains (equatorial, blue; intermediate, green; apical, red). **(D)** Movements of helices H, K and M sampled along the transition pathway. Three conformations are shown at decreasing transparency levels, starting from R" (lightest color), to T (darkest).

3.3.1.3 Sequence of events

A series of conformations visited along the transition pathway are displayed in Figure 10**C** for F_{min} = 0.5. We observe that, if the E domain (blue) is fixed as a reference, the A domain (red) closes down up and twists about the axial direction, and the intermediate (I) (green) domain moves away. Helix M (386-409) on the I domain, and helices H (233-244) and K (338-354) on the A domain are highlighted in Figure 10**D** (*left* diagram) and shown at three successive times from different perspectives (three diagrams on the *right*) to provide a clearer description of their movements. The three color shades, from light to dark, describe three snapshots along the reaction coordinate. The movement of the I domain, and the helix M in particular, seen on the *left* panel, will be shown below to be crucial in forming/breaking critical contacts during the allosteric transitions of the intact GroEL.

An interesting observation is the sequential order of events: first the E and I domains almost stick to each other and move coherently as a single rigid body, while the A domain undergoes an upward tilting and simultaneous twisting. These movements of the A domain are completed in the first half of the transition pathway from T to R". Then, slight rearrangements in the relative positions of the E and I domains occur, which expose the top portion of the A domain to bind the substrate and GroES. This sequence of events is consistent with the two-stage transition explored by the TMD simulations [33]. See the animations on http://www.ccbb.pitt.edu/People/yzheng/ for more details.

3.3.1.4 Which modes?

The number of modes involved in the transition between substates has been a question raised in a number of studies [160, 161]. Table 9 lists the number of modes, $m_T^{(k)}$ and $m_{R"}^{(k)}$, recruited at each

step as the molecule travels between the end points. An important observation is the increased involvement of higher frequency modes as we proceed away from the original state. In another words, the slow modes play an important role at the initial stages of deformation, and continue to play a role throughout the entire trajectory, although they are gradually complemented by increasingly larger subsets of higher frequency modes. The dominance of low frequency modes is consistent with the previously noted driving role of large scale motions in GroEL allostery [154].

Due to the limits of applicability of the harmonic potential, it is arguable how far away from the original energy minima we can extend the quadratic approximation. In the trajectory of $F_{min} = 0.5$, after 10 iterations (Table 9), the distance between the instantaneous intermediates ($\boldsymbol{d}^{(k)}$) falls from 12.33 Å to less than 4 Å. Only six low frequency modes out of the complete set of 1536 ANM modes accessible to the starting conformation (R") are recruited to reach this stage, along with seven slowest modes accessible to state T. Low frequency modes are known to be robustly determined by the overall shape and contact topology of the examined structure, i.e., they are insensitive the specific interactions and nonlinear effects, hence the broad use of ENMs in NMAs in recent years [15, 44]. This property emphasized in several reviews [14, 42, 43] is also supported by the close agreement between the low frequency modes from quasiharmonic analysis of MD trajectories and those from coarse-grained NMA; see, for example, references [162] and [163]. The fact that such a large portion (up to 2/3) of the reconfiguration occurs via these robust modes lends support to the applicability of *a*ANM. A small portion near the PTS needs, however, to be interpreted with care, as will be further analyzed below.

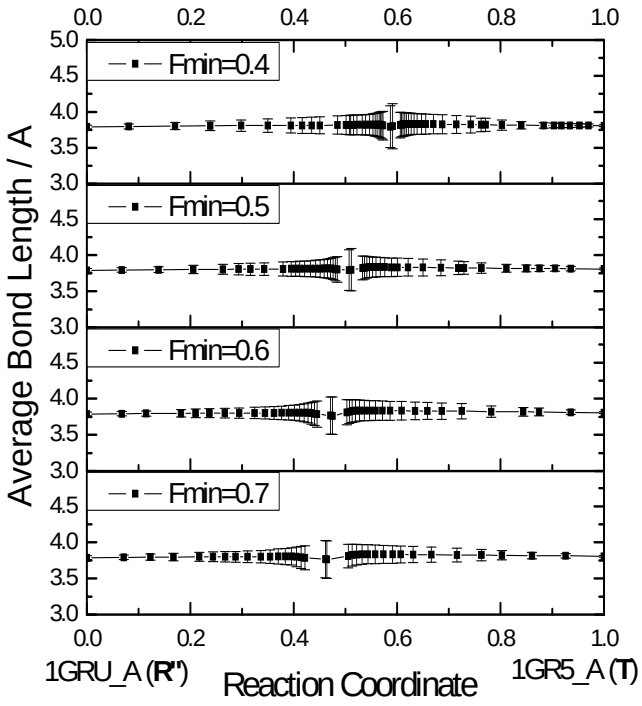

Figure 11. C^{α}-C^{α} bond lengths at intermediate states observed in *a*ANM analysis.

The error bars indicate the standard deviations from average value. The average bond lengths and their standard deviations for various F_{min} values with $f = 0.2$ are summarized in Table 12.The F_{min} in 0.5, 0.6 or 0.7 give almost the same average bond length, 3.81 ± 0.12 Å, which confirmed that the reconfigurations did not induce any unrealistic distortions in backbone bond lengths.

Table 12. Backbone bond lengths and shortest non-bonded distances in *a*ANM

F_{min}	Average backbone bond lengths and their standard deviation (Å)	Shortest non-bonded distances (Å)
0.4	3.82 ± 0.14	3.50
0.5	3.82 ± 0.12	3.43
0.6	3.81 ± 0.12	3.54
0.7	3.81 ± 0.12	3.55

In a strict sense, the normal modes provide information on the direction of motion near an energy minimum, and steps along these modes will be accurate to the extent that they are

infinitesimally small. In order to examine how the step size $s_A^{(k)}$ (or $s_B^{(k)}$) affects the course of structural reconfiguration, we repeated our calculations with various scaling factors f in the range $0.1 < f < 1$. As described above $f = 1$ allows for taking the full step size that minimizes $\boldsymbol{d}^{(k)}$ at a given iteration k, but entails possible unphysical distortions in the structure. Calculations with different f values showed that the parameter $f = 0.2$ can be safely adopted to avoid unrealistic deformations in backbone geometry. In particular, we monitored the C^{α}-positions and C^{α}-C^{α} bond lengths, which refer to bonded and non-bonded interactions in the coarse-grained model used here, to verify that the conservative steps generated with this scaling factor avoid steric clashes between residue pairs and unrealistic fluctuations in bond lengths. In Figure 11, the error bars indicate the standard deviations from average value. The average bond lengths and their standard deviations for various F_{min} values with $f = 0.2$ are summarized in Table 12. The F_{min} in 0.5, 0.6 or 0.7 give almost the same average bond length, 3.81 ± 0.12 Å, which confirmed that the reconfigurations did not induce any unrealistic distortions in backbone bond lengths.

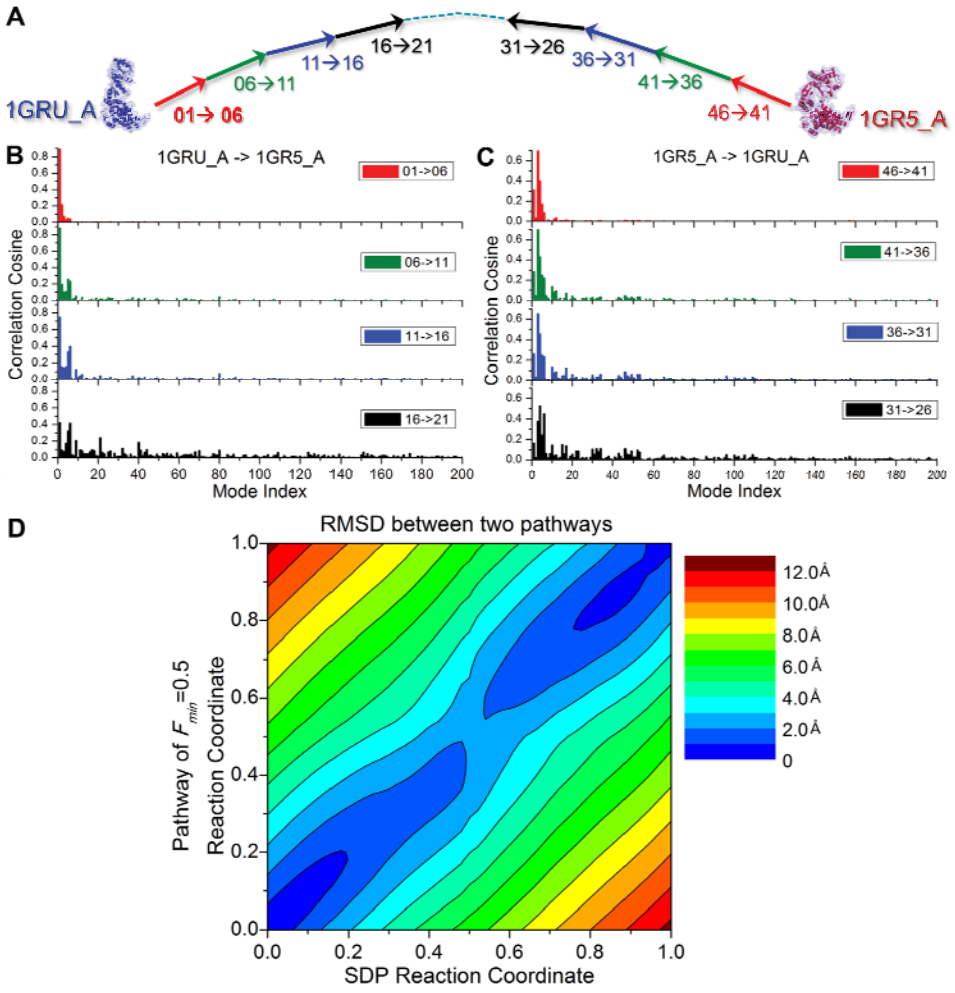

Figure 12. Comparison of *a*ANM results with steepest descent pathway (SDP).

(A) Fragmentation of the SDP pathway for the transition 1GRU ←→ 1GR5 of a subunit into nine macrosteps, consisting each of five frames. Same color scheme is adopted in panels B and C. **(B)** Correlation between SDP macrosteps and ANM modes accessible to the original conformation $\boldsymbol{R}_{R''}^{(0)}$. **(C)** Same as panel **B**, for the right portion of the trajectory, i.e. the reconfiguration from 1GR5_A to 1GRU_A using the eigenvectors $\boldsymbol{u}_i^{(1)}$ generated for 1GRU_A. Note that the early macrosteps from both directions are accounted for by a few slowest ANM modes, while increasingly higher modes are being recruited as the molecule proceeds away from its original conformation, consistent with the results found by *a*ANM (see Table 9). **(D)** RMSD values between the intermediate conformations sampled by the *a*ANM and SDP methods. The *a*ANM results refer to the trajectory F_{min} = 0.5. The RMSDs between pairs of intermediates remain lower than 2.0 Å at all steps (see the color-coded scale on the right).

3.3.1.5 Comparison with action minimization results

Toward assessing the consistency of our results with those predicted by other analytical methods, we examined the transition pathway predicted for the same model by the action minimization algorithm [59, 62]. The resulting trajectory, called the steepest descent pathway, led to the energy profile shown by the black curve in Figure 10**B**. The barrier height is slightly lower than that obtained by *a*ANM, while the shape of the energy profile shows close resemblance: a peak around $x(n) = \sim 0.52$, preceded by minimal energy cost for an extended portion of the reaction coordinate, followed by a sharp increase near the peak, and then a smooth decrease after the peak.

Toward a more critical analysis of the modes that contribute to the SDP, we reorganized the SDP trajectory (consisting of 46 frames) into a series of $k = 9$ (*macro*)*steps* by collapsing each set of five consecutive frames into a macrostep (Figure 12**A)** and we calculated the deformation vector $\Delta \boldsymbol{R}_k^{\mathrm{SDP}} = \boldsymbol{R}_{n+5}^{\mathrm{SDP}} - \boldsymbol{R}_n^{\mathrm{SDP}}$ for each macrostep. The following questions were raised: Which ANM modes effectively contribute to these macrosteps? Do SDP macrosteps exhibit the same tendency as *a*ANM to originally proceed via softer modes and gradually recruit increasingly larger subsets of modes? How similar are the conformations visited along the *a*ANM and the SDP? To this aim, we evaluated the correlation cosine between $\Delta \boldsymbol{R}_k^{\mathrm{SDP}}$ and the ANM modes $\boldsymbol{u}_{iA}^{(1)}$ and $\boldsymbol{u}_{iB}^{(1)}$ accessible to original states $\boldsymbol{R}_A^{(0)}$ and $\boldsymbol{R}_B^{(0)}$. The results are shown as a function of mode index i in the respective panels **B** and **C** of Figure 12. The correlation cosines represent the relative contributions of the intrinsically accessible ANM modes to the SDP macrosteps. In accord with the results from *a*ANM, only very few modes at the low frequency end of the spectrum contribute to the SDP macrosteps in the close neighborhood of the original states (red plots). The slow modes contribute by almost by the same amount as those observed in *a*ANM at the successive stages of the transition pathway. The contribution of higher

frequency modes, which is negligibly small at early stages, gradually increases, consistent with the *a*ANM.

An even more direct comparison of the conformations sampled by the *a*ANM and the SDP is provided by the map in Figure 12**D**. The map displays the RMSDs between the conformations sampled by the SDP at successive steps (abscissa) and those sampled by the *a*ANM algorithm (ordinate) using $F_{min} = 0.5$. The RMSDs remain lower than 2 Å for the most part of the trajectory.

3.3.2 Allosteric transitions of the intact chaperonin

3.3.2.1 Intrinsic dynamics of the T/T, R/T, R'/T and R"/R states

Calculations performed above for a single subunit demonstrated that the transition pathways between the functional forms of the single subunit can be delineated by the *a*ANM algorithm. In particular, Table 11 showed that a few low frequency modes can account for a significant portion of the structural changes that take place between the functional forms. We now explore the allosteric transitions in the intact chaperonin. We begin by examining if/how the low frequency modes play equally an important role in the allosteric dynamics of the intact GroEL.

The counterparts of Table 11 for the intact chaperonin are Table 13. Here, the major difference is the fact that the changes occur via non-degenerate modes that maintain the heptameric symmetry of the GroEL structure, similar to the previously noted [34, 36] dominant role of non-degenerate ANM modes in enabling the maturation of icosahedrally symmetric viral capsids. Results are listed for the first 1, 3 and 6 non-degenerate modes. Note that the set of ANM modes accessible to the GroEL complex is larger than that of the single subunit by a factor of 14. We might thus expect to recruit a larger number of modes in the low frequency regime to

achieve the same fractional contribution to observed changes. However, it is again striking to observe that six non-degenerate modes are sufficient to generate intermediates that are almost half way through the transition between the end points. The original RMSD values of 6.72, 10.85 and 11.54 Å between the end states of the three respective steps T/T → R/T, R/T → R'/T and R"/R → T/R are reduced to 3.26, 6.35 and 7.87 Å upon moving along these modes (see Table 13).

Table 13. The contribution of lowest frequency modes to the three major steps of the chaperonin allosteric cycle[(a)]

Step	Starting struc.	Target struc.	Original RMSD (Å)	RMSD by the *1st* non-degenerate mode			RMSD by the *first 3* non-degenerate modes			RMSD by the *first 6* non-degenerate mode		
				RMSD (Å)	F_{min}	mode	*RMSD* (Å)	F_{min}	modes	*RMSD* (Å)	F_{min}	modes
T/T to R/T	1GR5	2C7E	**6.72**	**5.54**	0.32	1	4.91	0.46	1+10+ 11	3.97	0.65	1+10+11+ 28+51+61
	2C7E	1GR5		5.83	0.24	3	4.66	0.52	3+12+20	3.42	0.77	3+12+20+ 35+49+56
	Deform simultaneously			5.54			4.65			**3.26**		
R/T to R'/T	2C7E	2C7C	**10.85**	10.34	0.09	3	9.96	0.15	3+17+20	6.35	0.66	3+17+20+ 29+44+49
	2C7C	2C7E		10.68	0.03	3	9.96	0.16	3+4+10	9.34	0.27	3+4+10+ 20+21+22
	Deform simultaneously			10.34			9.77			**6.35**		
R"/R to T/R	1GRU	2C7E'[(b)]	**11.54**	11.22	0.05	3	9.35	0.34	3+10+17	8.34	0.48	3+10+17+ 20+22+30
	2C7E'	1GRU		11.19	0.06	3	9.33	0.35	3+20+27	8.52	0.46	3+20+27+ 29+32+35
	Deform simultaneously			11.19			8.43			**7.87**		

[(a)] Major steps are those involving an RMSD larger than 3.2 Å (~resolution of some structures) between the end points.
[(b)] 2C7E' is same as the PDB structure 2C7E, except for the inversion shown in step **F** of the allosteric cycle in Figure 7.

3.3.2.2 Energy profile and operating modes

We now proceed to the *a*ANM study of the transition T/T → R/T → R"/R. These three states are selected because they differ by an RMSD of 6.72 Å and 11.03 Å, respectively (see Table 10

part **B**). R'/T and R"/T differ from R"/R by less than 3 Å, which is comparable to the resolution of the structures, hence their exclusion from the above scheme. The results are presented in Figure 13 and Table 14**.** The energy profile along the transition R/T → R"/R (Figure 13**A)** exhibits a shape similar to that obtained with the single subunit; the PTS is closer to the R"/R state ($x(n)$ = 0.58), and the barrier is much lower than that encountered upon interpolation (orange curve) between the end points.

Figure 13**B** displays the histograms of modes contributing at various iterations (k = 1, 6, 11 and 15), starting from the R/T state. The tendency to recruit higher modes as the PTS is approached is clearly seen, consistent with the trend observed in the single subunit with both *a*ANM and SDP. Table 14 shows that the RMSD between the intermediate conformations reduces to 1.93 Å by selecting modes from amongst the slowest 636 modes. More detailed examination shows that among them 97 non-degenerate modes contribute to 90% of the structural change. This subset amounts to less than 0.5 % of the spectrum of ANM modes (~2.2 x 10^4 of them).

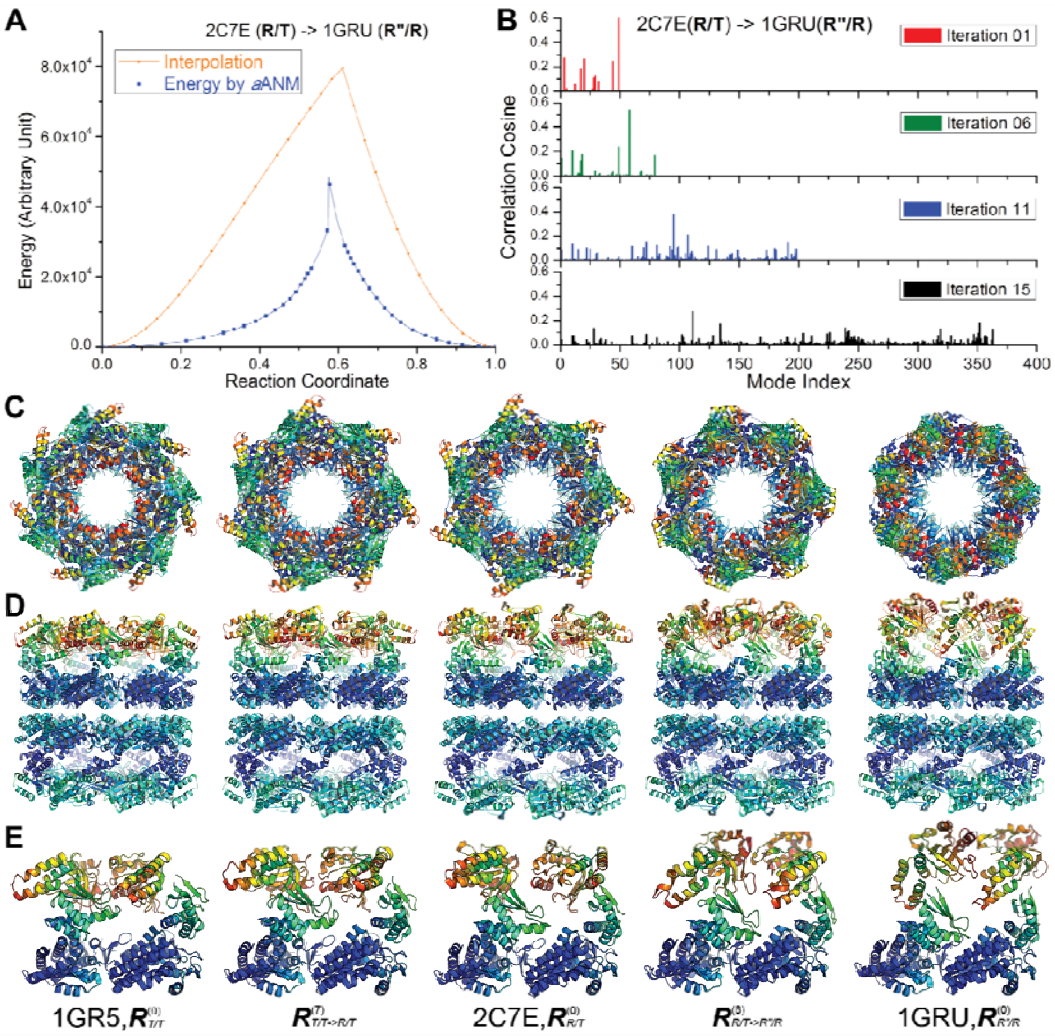

Figure 13. Transitions between T/T → R/T → R"/R forms for intact GroEL complex.

(A) Energy profiles of the intact GroEL complex (R/T → R"/R) along the reaction coordinate computed by ***a*ANM** (blue) and Cartesian interpolation (orange) using the double-well potential given by Eq. (3.7). **(B)** Contribution of different modes at various steps (1, 6, 11 and 15) along the transition R/T → R"/R. Broader numbers of higher frequency modes are recruited as the structure approaches the energy barrier (see Table 14). **(C)** Top view of structures sampled along the transition. Snapshots corresponding to conformations $\boldsymbol{R}^{(0)}_{T/T}$, $\boldsymbol{R}^{(16)}_{T/T->R/T}$, $\boldsymbol{R}^{(0)}_{R/T}$, $\boldsymbol{R}^{(18)}_{R/T->R"/R}$ and $\boldsymbol{R}^{(0)}_{R"/R}$ are shown. **(D)** Side view of the same structures. **(E)** Close-up views of pairs of adjacent subunits. The diagrams in panels **C-E** are color-coded according to the mobilities of residues (red: most mobile; blue: almost fixed). Note that the equatorial domains of *cis* ring subunits are almost fixed, while the largest motions occur at the apical domains of the same subunits.

Panels **C, D** and **E** display the evolution of the intact structure viewed from the top (**C**), and from the side (**D**), along with a close-up view of two adjacent subunits on the *cis* ring (**E**). See http://www.ccbb.pitt.edu/People/yzheng/ for movies. The *trans* ring is generally observed to undergo moderate changes between the T to R states. The *cis* ring, on the other hand, undergoes concerted rotations and extensions at the apical and intermediate domains. Notably, the intermediate domains move toward the cleft between neighboring equatorial domains, while the apical domains extend along the vertical (cylindrical axis) direction accompanied by a rotation about the same axis. These motions result in the enlargement of the central cavity along with the exposure of the apical domains on the *cis* ring to bind the GroES-binding flexible loop K15-T36.

Table 14. *a*ANM data for the transition of intact GroEL complex(*)

	1GR5 (**T/T**) to 2C7E (**R/T**)			2C7E (**R/T**) to 1GRU (**R"/R**)		
k	$m_{T/T}^{(k)}$	$m_{R/T}^{(k)}$	*RMSD* (Å)	$m_{R/T}^{(k)}$	$m_{R"/R}^{(k)}$	*RMSD* (Å)
0	--	--	6.72	--	--	11.03
1	51	20	6.17	49	121	9.70
2	51	35	5.67	49	122	8.53
3	61	51	5.04	49	139	7.58
4	89	51	4.55	49	142	6.86
5	94	65	4.11	54	181	6.14
6	118	104	3.74	79	188	5.58
7	218	118	3.40	120	189	5.07
8	238	183	3.10	95	209	4.61
9	282	205	2.80	86	310	4.21
10	326	229	2.55	142	338	3.80
11	425	282	2.32	198	384	3.44
12	515	346	2.14	240	471	3.10
13	620	510	1.95	318	576	2.80
14	942	616	1.80	321	731	2.53
15	-	-	-	363	1054	2.31
16	-	-	-	519	1380	2.13
17	--	--	--	636	1869	1.93
18	--	--	--	792	2281	1.23

(*) $k = 0$ refers to the original RMSD between the end points. Results are obtained with $F_{min} = 0.5$.

3.3.2.3 Redistribution of inter-subunit salt bridges and implications on intra-ring cooperativity

Salt bridges at the interface between the *cis* and *trans* rings have been pointed out in site-directed mutagenesis experiments by Saibil and coworkers to play a critical role in communicating allosteric signals between the two rings and the co-chaperonin binding site [151]. The redistribution of critical interactions accompanying the global changes in structure has been proposed to be a possible mechanism for regulating allosteric signal transduction [13]. Here we examine more closely the changes in the distances between salt-bridge forming residues during

the transitions of the intact chaperonin, and discuss our results in relation to previous experimental and computational data.

Experiments conducted with Arg197Ala mutant [164] pointed to the functional role of the salt bridge between E386 and R197 on adjacent subunits in the *cis* ring. This inter-subunit salt-bridge was also noted in early cryo-EM studies [142]. It forms in the T state of the ring (1GR5) (Figure 13 **A** and **B**), and it has been proposed to be an essential component of the positive intra-ring cooperativity [165]. The R" state of the same ring (1GRU) indicates, on the other hand, a new salt bridge, formed between E386 and K80 (on the E domain of the neighboring subunit) (Figure 13**D**). The *a*ANM results shed light to the mechanism of this interchange of salt bridges. The separation between E386 and R197 α-carbons, originally equal to 15.7 Å (their charged ends being separated by 3.2 Å), gradually increases by the downwards motion of the intermediate domain. After 18 iterations (for T/T→R/T), the distance between E386 and R197 becomes larger than that between E386 and K80, and K80 replaces R197 to form a salt bridge with E386 (Figure 13**C**). Afterwards, R197 moves dramatically away from E386, led by the opening of the apical domain. Meanwhile the downward movement of helix M continues until the distance between the C^{α}-atoms of the new salt-bridge-forming residues E386 and K80 reduces to 8.6 Å.

These intra-ring rearrangements are in accord with previous TMD simulations [33]. However, TMD runs showed heterogeneous subunit behavior, while the *a*ANM trajectory retains the sevenfold rotational symmetry. In each ring, the movements of all subunits are identical and coherent (Figure 13**C**). In Figure 13**B** we noted that not all slow modes in the lower end of the spectrum contribute to the transition. Instead a subset inducing cylindrically symmetric changes is recruited. For example, in the first iteration, while 51 modes are recruited, the contributions of

modes 2 through 9 are almost zero, and we have large contributions from the non-degenerate modes 1, 10, 11, 28 and 51 (Table 13). Our results support the Monod-Wyman-Changeux (MWC) view [10, 166] of cooperative, symmetric (non-degenerate) modes of motion controlling allostery. Likewise, GroEL engineering experiments provide evidence for the concerted nature of the allosteric transition [165].

3.3.2.4 Comparison with the results from recent Brownian dynamics simulations

A recent examination of the time evolution of salt-bridges formed/broken at the steps T → R and R → R" of the GroEL allosteric cycle brought attention to a number of key interactions [154]. Those results have been obtained by Hyeon, Lorimer and Thirumalai, by performing BD simulations, for a dimer, using state-specific Go-like potentials for the end points [154]. Therefore, they differ from ours in their method, model and parameters. Yet, in order to see if similar trends could be detected in spite of the differences in the methodology, we undertook a careful examination of the formation/breakage of salt bridges along the *a*ANM predicted reaction coordinates.

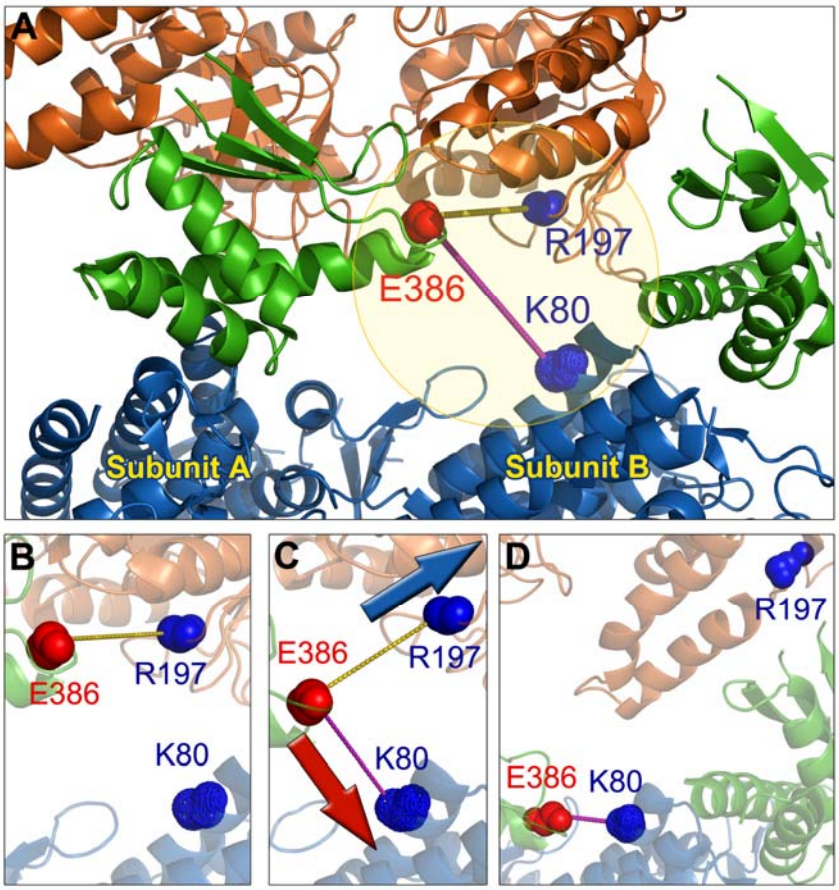

Figure 14. *cis* ring inter-subunit interactions during the transition T → R".

(A) Intersubunit interface near the intermediate domains (green) of two adjacent subunits in the *cis* ring. The backbones are shown in cartoon view and colored by domains: A (orange), I (green), and E (blue). Backbone atoms of three charged residues are shown by spheres. Positively and negatively charged residues are colored blue and red, respectively. **(B)** The inter-subunit hydrogen bond, E386-R197, in the T state of the *cis* ring (1GR5). **(C)** During the transition to state R"/R, residue E386 in the I domain moves towards K80 (blue sphere) in the E domain of the adjacent subunit, while R197 on the A domain moves away from E386. **(D)** The final configuration in the R" state of the *cis* ring, represented by 1GRU. Residue E386 now forms a new hydrogen bond with K80.

In the interest of making an even more stringent, quantitative comparison between the two sets of data, we have considered the relative rates of formation/rupture of different salt bridges. Our approach does not contain 'time' explicitly. Instead we have the 'number of

iterations'. In order to evaluate the relative time dependence of the salt bridge formation/breakage events, we normalized one of our curves, called 'reference curve', with respect to its counterpart obtained [154] by Hyeon *et al.* (2006). This permitted us to plot the accompanying time dependence of all other salt bridges as a function of time (Figure 15) and determine the best fitting kinetic data. The results, compiled in Table 15 exhibit striking quantitative agreements between the *a*ANM results and those from BD simulations.

Figure 15 displays our results for the respective T→R (panel **A**) and R→ R" (panel **B**) transitions of *cis* ring subunits obtained by aANM for the intact chaperonin. These are the C^{α}-C^{α} distances between intra- or intersubunit salt-bridge forming residues, plotted as a function of reaction coordinate. Interestingly, particular pairs or residues exhibit a smooth increase/decrease in their distance, whereas others are more sluggish. The former group of pairs which readily associate/dissociate may be viewed as driving interactions. We notice, for example, the smooth decrease in K80-D359 distance along with increase in D83-K327 distance (panel **B**), indicative of the compensating formation and breakage of these two intra-subunit salt bridges prompted at an early stage of the R → R" transition. These changes undergone early on are in sharp contrast to the salt bridges involving E257 and E386, which exhibit strong resistance to dissociate or associate (panels **A** and **B**). The leading role of K80-D359 and D83-K327 is in agreement with the mechanism observed in BD simulations [154]. The formation of the K80-D359 salt-bridge has been pointed out therein to be the major driving force for the R → R" transition [154]. Also, this formation has been pointed out to be coupled to the dissociation of D83-K327 [154]. Other similar features include the fast increase in R58-E209 and decrease in P33-N153 distances at short times during T → R, in contrast to the slower responses of D83-K327 and E257-R268 at the same stages (panel **A**), succeeded by the high stability of P33-N153 during R → R", the

resistance of E257-K321 and E257-R322 to come closer at the initiation of T→R, in contrast to their predisposition to dissociate at the onset of R → R". Finally, it is interesting to notice that the non-monotonic behavior of the pair E257-R268 in panel **B** conforms to previously observed [154] ensemble-averaged behavior for the same pair; while that of R197-E386 observed here departs from the smooth increase previously reported.

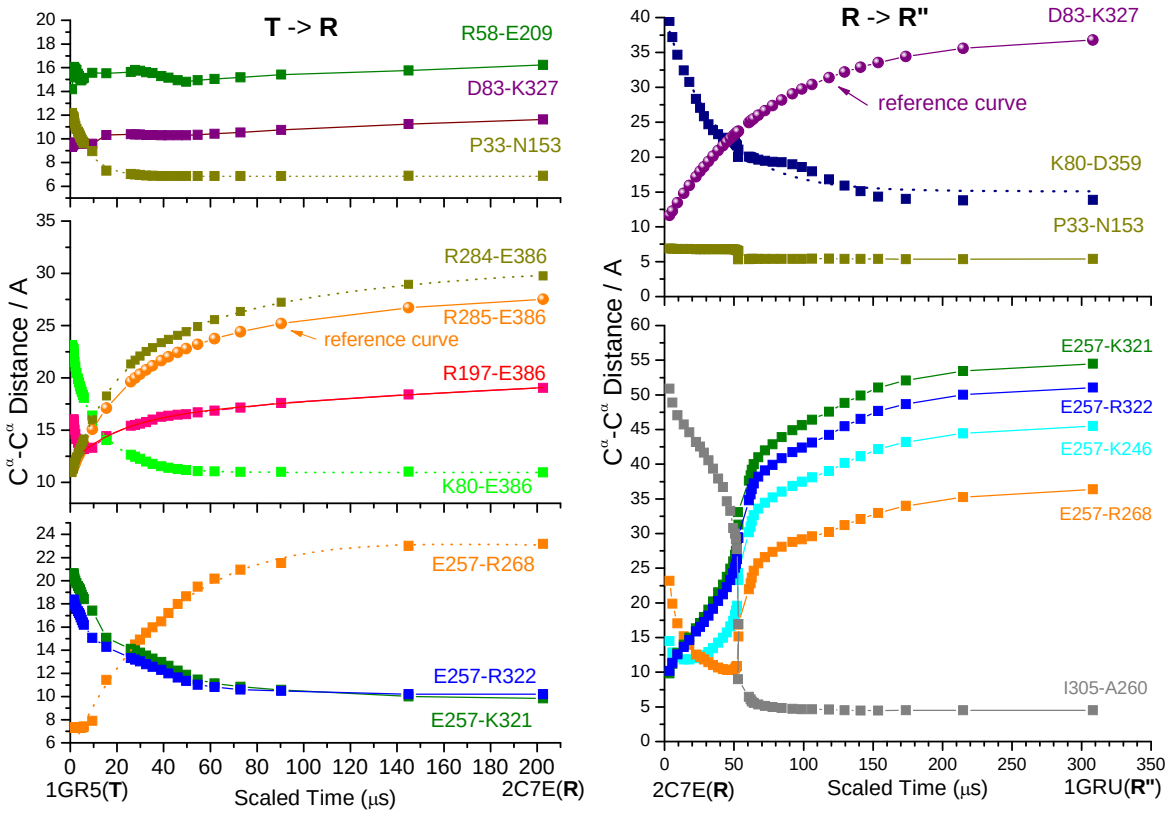

Figure 15. Changes in the distances between salt-bridge forming pairs during transition.

Results are shown for **(A)** T→ R and **(B)** R→ R" transitions. The time dependence is deduced by normalizing one of the curves in each panel, labeled 'reference curve', with respect to those presented in the Figure 10 and Figure 12 of ref. [154]. See Table 15 for the kinetic expressions and the comparison with the results from BD simulations by Hyeon *et al*[154].

Table 15. Comparison of the kinetics of salt-bridge forming pairs obtained by *a*ANM and BD simulations[(a)]

Transition	Salt Bridge	Time dependence of C^{α}-C^{α} distance (*a*ANM) [(b)]	R^2 (*a*ANM)	<d(t)> (from reference[154])	Correlation coefficient [(c)]
T → R	D83-K327	$11.8 - 2.3e^{-t/101.1}$	0.91	$13.4 - 4.9\ e^{-t/100.0}$	0.999
	P33-N153	$6.8 + 5.8\ e^{-t/7.7}$	0.99	$7.3 + 4.2e^{-t/6.3}$	0.998
	R284-E386	$30.7\text{-}9.9e^{-t/15.8} - 10.4e^{-t/84.2}$	0.99	$29.7 - 8.1e^{-t/20.8} - 8.4e^{-t/85.8}$	0.999
	R285-E386	Reference curve	--	$28.4 - 6.6e^{-t/19.1} - 8.1e^{-t/88.8}$	1
	R197-E386	$22.4\text{-}4.1e^{-t/20.0} - 6.4\ e^{-t/307.2}$	0.99	$20.9 - 2.6e^{-t/0.67} - 6.4e^{-t/96.7}$	0.975
	K80-E386	$11.0+10.1e^{-t/14.1} + 4.0e^{-t/2.5}$	0.99	$10.4 + 7.6e^{-t/12.1} + 2.2e^{-t/61.8}$	0.993
	E257-R268	$22.5\text{-}169.2e^{-t/58.3}+151.9e^{-t/62.0}$	0.99	$21.8 - 4.2e^{-t/26.2} - 7.9e^{-t/66.4}$	0.998
	E257-R322	$10.3 + 8.1e^{-t/24.8}$	0.99	N/A	--
	E257-K321	$10.1+ 10.7e^{-t/26.6}$	0.99	N/A	--
R → R"	D83-K327	Reference curve	--	$37.3 - 26.9\ e^{-t/77.9}$	1
	E257-K321	$58.1 - 55.1e^{-t/73.1}$	0.94	N/A	--
	E257-R322	$54.6 - 50.7e^{-t/76.2}$	0.94	N/A	--
	E257-K246	$51.6 - 47.6\ e^{-t/95.8}$	0.86	N/A	--
	K80-D359	$15.1 + 25.2e^{-t/37.4}$	0.98	$14.1+ 26.4e^{-t/28.0}$	0.991

(a) BD simulations results were reported by Hyeon *et al.* (2006) [154]; time (t) in microseconds.

(b) *a*ANM results are reported for all salt-bridges that exhibited a monotonic time dependence (single or double exponential) with $R^2 > 0.85$.

(c) between the two sets of data (*a*ANM and BD) for each salt bridge.

3.3.2.5 Critical interactions formed/broken at the transition involve conserved residues

The above results (Figure 15) reveal that certain interactions play a key role in stabilizing particular states, once formed. For example, the inter-subunit E257-R268 salt bridge appears to be crucial for the T state of the *cis* ring; whereas the inter-subunit hydrophobic contact I305-A260 seems critical to stabilizing the R" state. These interactions are usually disrupted only at the PTS.

Toward a more systematic assessment of such critical interactions, we focused on the changes in native contacts occurring at the PTSs. A set of 2028 native contacts are shared

between the R and R" states (native contacts being defined as C^{α}-C^{α} distances lower than 7 Å, in the present model) while the two respective states exhibit 137 and 105 distinctive native contacts. The questions are: How does the number of native contacts vary along the transition pathway? Which contacts among those differentiating the two end structures are maintained along the trajectory and which ones are broken or formed only at the transition state?

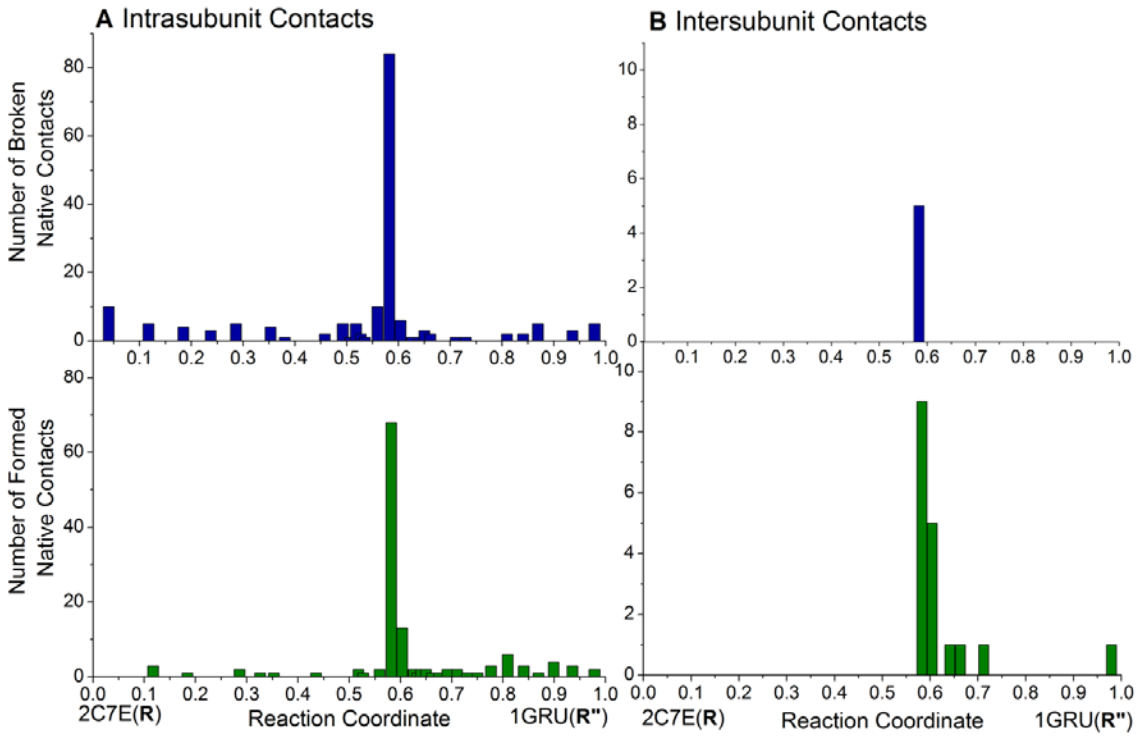

Figure 16. Evolution of native contacts along the structural transition from R to R".

The number of intra-subunit (panel **A**) and inter-subunit (panel **B**) native contacts that are disrupted (upper panel) and formed (lower panel) *vs.* the reaction coordinate. The results refer to $F_{min} = 0.5$ for *cis* ring subunits along the transition from 2C7E (R/T) to 1GRU (R"/R). Each bar represents the number of native contacts formed/broken at a given *a*ANM iteration. Note the sharp increase near the energy barrier. See **Figure 17** for the corresponding critical contacts.

Figure 16**A** displays the time evolution of native contacts observed for a single subunit during the transition R → R"; and panel **B** shows its counterpart for inter-subunit contacts observed in the complex. Clearly, most of the native contacts remain unchanged throughout a large portion of the trajectory, while significant changes occur near the PTS, both with regard to the disruption (*top* panel) and formation (*bottom* panel) of contacts. Trajectories generated by varying *a*ANM model parameters in the ranges $0.4 < F_{min} \leq 0.7$ and $0.2 \leq F_{min} \leq 0.5$ yielded almost identical results, confirming the strong preference of the molecule to redistribute native contacts only in the vicinity of the PTS, while the large portions of the trajectories are enabled by minimal changes in native contacts.

A closer examination reveals the residues involved in critical interactions tend to be strongly, if not fully, conserved. Table 16 lists the inter-subunit residue pairs broken/formed at the transition, also called 'critical intra-ring contacts', along with their conservation scores evaluated using ConSurf [167, 168]. The E domains make indeed many more inter-subunit contacts than the A domains (Figure 17). As a result the A domains can move independently of one another, while the E domains are constrained [154], which explains their breakage at the critical state only. It is also interesting to note that the only I domain residues that make critical contacts (Ala384 and Thr385 next to the N-terminus of the M helix) undergo large reorientations (Figure 10**D**). We also note the segment A384-V387 that has been pointed to play an important role in mediating positive intra-ring cooperativity [27].

Table 16. Critical inter-subunit contacts broken/formed during the transition R → R"

	Chain A			Chain B		
	Residue type	**Index**[(a)]	**ConSurf score**[(b)]	**Residue type**	**Index**[(a)]	**ConSurf score**[(b)]
Breaking	VAL	39 (E1)	0	**GLU**	**518** (E2)	9
	ASP	**41** (E1)	9	**MET**	**520** (E2)	9
	GLY	**45** (E1)	9	**GLN**	**72** (E1)	9
	ALA	46 (E1)	0	**MET**	**69** (E1)	9
	ALA	46 (E1)	0	**GLU**	**76** (E1)	9
Forming	**ALA**	**26** (E1)	9	**CYS**	**519** (E2)	9
	VAL	**29** (E1)	9	**GLU**	**518** (E2)	9
	LYS	34 (E1)	0	MET	114 (E1)	9
	ARG	**36** (E1)	9	**THR**	**517** (E2)	9
	ASN	**37** (E1)	9	**CYS**	**519** (E2)	9
	LEU	**40** (E1)	9	**THR**	**522** (E2)	9
	PRO	**47** (E1)	9	**GLN**	**72** (E1)	9
	ILE	60 (E1)	7	**LYS**	**4** (E1)	9
	GLU	**61** (E1)	9	ALA	2 (E1)	0
	THR	210 (A)	1	GLN	351 (A)	0
	ALA	**260** (A)	9	**GLU**	**304** (A)	9
	ALA	**260** (A)	9	ILE	305 (A)	0
	ALA	**260** (A)	9	GLY	306 (A)	0
	VAL	**264** (A)	9	GLY	306 (A)	0
	ALA	**384** (I2)	9	TYR	506 (E2)	0
	ALA	**384** (I2)	9	**SER**	**509** (E2)	9
	THR	**385** (I2)	9	TYR	506 (E2)	0
	THR	**385** (I2)	9	**SER**	**509** (E2)	9

[(a)] Conserved residues are highlighted in boldface. The domains are shown in parentheses. The equatorial domain residues are Met1-Pro137 (E1) and Val411-Pro525 (E2); intermediate domain residues are Cys138-Gly192 (I1) and Val376-Gly419(I2); and the apical domain (A) consists of a contiguous segment Met193-Gly375.

[(b)] The scores are based ConSurf Server [167, 168] calculation for GroEL sequence.

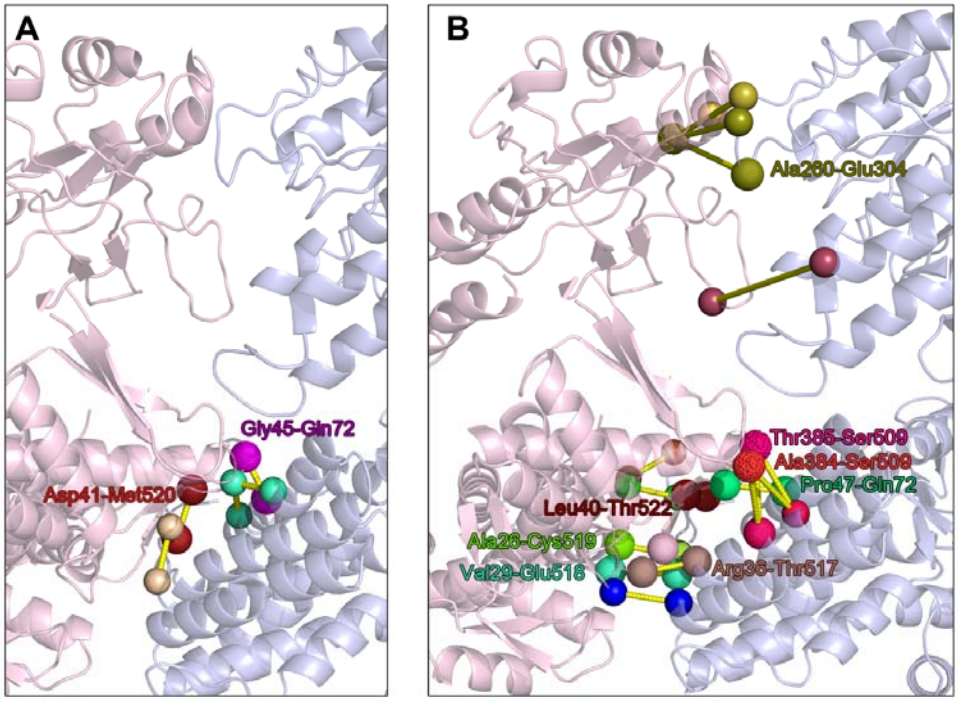

Figure 17. Redistribution of inter-residue contacts at the transition from R to R" state.

Panel (**A**) shows the inter-residue contacts between adjacent subunits of the *cis* ring, which break up during the transition. The two subunits are colored in light pink and blue. Contact pairs are represented by spheres at the C^{α} position, and by distinctive colors. Similarly, panel (**B**) shows the contacts newly formed during the transition. See Table 16 for the complete list of residue pairs shown here. The contacts involving a pair of conserved residues are labeled using the same colors as the corresponding residues.

Broken pairs distinguished by their high conservation are Asp41-Met520 and Gly45-Gln72; and those conserved contacts made at the PTS are Ala26-Cys519, Val29-Glu518, Arg36-Thr517, Asn37-Cys519, Leu40-Thr522, Pro47-Gln72, Ala260-Glu304, Ala384-Ser509 and Thr385-Ser509 (Figure 17). Notably, some of these residues have been already reported to be functional. The mutation Gly45Glu, for example, causes defective release of partial folded polypeptide without reducing ATPase or GroES binding [169, 170]. Val264Ser reduces GroES and polypeptide binding affinity [169]; Ala260 is involved in binding and unfolding rhodanese,

as shown in MD simulations [171]. Mutants Arg36His, Glu76Lys, Val264Ile, Thr385Ile, Met520Ile and Thr522Ile give rise to active single-ring GroEL [172]. Finally, we note that all four residues, Val39, Asp41, Gly45 and Ala46, listed in Table 16 to break at the energy barrier, belong to the segment Val38-Ile49, pointed out in our previous study [27] to play a key role in establishing the intra-ring allosteric communication.

3.4 DISCUSSION

GroEL is a molecular machine that has been broadly studied in recent years using both experimental and theoretical or computational methods. Yet, a structure-based analysis of the transition of the intact chaperonin between its functional forms has been held back by the large size of the chaperonin. The *a*ANM method is proposed as a first approximation toward approaching this challenging task.

The conformational transition results from the competition of two counter effects at a minimum, one intramolecular, and the other intermolecular: (1) any deformation away from the original equilibrium states involves an uphill motion along the energy surface, but the lowest frequency motions incur the lowest ascent in energy (intramolecular); (2) longer pathways experience higher frictional resistance from the surroundings, such that the shortest pathway may be favored (intermolecular) at the expense of internal strains. In the *a*ANM model, one can balance these two counter effects by adopting different threshold correlation cosine (F_{min}), for different proteins in various conditions. Higher F_{min} values emphasize the effect (2), with the upper limit $F_{min} = 1$ restricting the pathway to a pure interpolation. Lower F_{min} values on the

other hand allow for occasional departures from the targeted direction, with the limit $F_{min} = 0$ tolerating steps orthogonal to the target direction.

Central to the development of this methodology were approximations and parameters that needed to be critically tested, hence our extensive examination of the transitions of a GroEL subunit and comparison with the results from other methods, including SDP (action minimization), MD simulations [33] and experimental studies. Multiple pathways were explored to this aim by varying the parameters F_{min} and f, which control the respective direction and size of the *a*ANM steps. It is conceivable that multiple routes are accessible between the two endpoints, given the multidimensional character of the energy landscape. For example, broad transition state ensembles have been observed by Hyeon, Lorimer and Thirumalai in their simulations of the GroEL allosteric transitions using a self-organized polymer model [154]. However, one or more are more probable, and *a*ANM method by definition aims at sampling such structurally favored paths. Of interest was to identify the common features, if any, of the paths sampled with different parameters, and to identify among them those most closely agreeing with the data from other methods and experiments. Notably, the generated pathways invariably exhibited the following common features, observed for a single subunit, and confirmed in the intact chaperonin:

(i) The recruitment of low frequency modes near the original state, succeeded by the recruitment of higher frequency modes near the barrier (see for example Table 9, Figure 9, Figure 12**A-C** and Figure 13**B**),

(ii) The dominant role of a few well-defined low frequency modes for achieving substantial (at least halfway) displacement in the 'functional' direction (Table 10 and Table 13**)**,

(iii) The close agreement between pathways and energy profiles generated with different parameters, observed over a large portion ($x(n) \leq 0.4$ and $x(n) \geq 0.7$) of the reaction coordinate, also in accord with SDP results (see for example Figure 10**A-B**, Figure 12**D**), consistent with the predominance of well-defined modes at the initial stages of the structural changes,

(iv) Barrier heights significantly lower than that incurred upon linear interpolation of coordinates between the two endpoints (Figure 9**B**, Figure 12**A**),

(v) Protection of the native state inter-residue contacts throughout a surprisingly large portion of the transition pathway, and formation/breakage of native contacts, mostly at the E domains, only in the close neighborhood of the PTS (Figure 17 and Table 16).

Having established these common features, a closer analysis of intermediate conformations, energetics and mode distributions in comparison to those from the rigorous SDP method supported the use of $F_{min} = 0.5$ as an optimal parameter for further calculations for the intact chaperonin. A discriminative feature that supported this parameter was the position of the energy barrier along the reaction coordinate, in accord with SDP. The close neighborhood of the energy barrier is indeed the region most sensitive to the choice of parameters. Notably, this region involves local rearrangements (RMSD < 2 Å) comparable (or below) the resolution of examined structures. Therefore the results in this regime should be interpreted with caution. However, the set of 'critical' contacts identified for the R → R" transition (listed in Table 16) have been verified to be robustly reproduced over a relatively broad range of *a*ANM parameters ($0.4 < F_{min} \leq 0.7$ and $0.2 \leq f \leq 0.5$). Another observation in strong support of *a*ANM results with $F_{min} = 0.5$ is the striking qualitative and quantitative agreement between the dynamics of salt-

bridges observed here and those found by BD simulations with state-dependent energy functions and parameters [154], presented in Figure 15, and Table 15.

The application of *a*ANM to GroEL therefore elucidates not only highly probable pathways or the hierarchic contribution of modes to achieve the transition; but it also provides insights into key interactions that initiate the transition (e.g., formation of the K80-D359 salt-bridge), or those that form/break at the transition state(s). Such inter-subunit contacts disrupted and formed at the PTS are illustrated at the respective panels **A** and **B** of Figure 17. Notably, the majority of the residues involved in these critical interactions are highly conserved (Table 16), and some of them have been observed in previous site-directed mutagenesis experiments to affect GroEL machinery [169-172]. The importance of the other identified critical contacts to the functional transitions of the chaperonin remains to be further established by experiments.

On a practical side, the major utility of the method lies in its application to the transitions of supramolecular systems beyond the range of exploration of other computational methods. The computing time in the present method is several orders of magnitude shorter than that required in regular MD or BD simulations. Here, sampling the transition pathway of the intact GroEL (of approximately 8,000 residues) takes less than 255 min CPU time (or about 33 hr real time) on a 2G Hz Linux server.

Finally, from a broader perspective, it is worth noting that this type of calculation is possible only to the extent that the conformational changes intrinsically favored by the overall architecture comply with those required to achieve the biological function (or the allosteric cycle, here). In principle, the configurations are defined in a *3N*-dimensional space, defined by *3N*-coordinates (provided that we adopt a coarse-grained description of *N* position vectors, one per residue). Here, we let the molecule move in the subspace of 5-6 coordinates only predicted by

the ANM, and we can see that more than ½ of the trajectory along the conformational space is traversed by the molecule by simply moving along these soft coordinates. High frequency modes need not be recruited until the energy barrier is approached. The conformational changes are therefore effectuated to a large extent through low frequency modes, which are the least expensive (from energetic point of view), or the most favorable (from entropic point of view) modes among a multitude (3*N-6* of them) of possible modes/directions of reconfiguration. This brings up the issue of a possible evolutionary selection of 3-dimensional structures (or inter-residue contact topologies) whose energy landscape is most conducive to functional changes in conformation. The selection of conformational changes entropically favored by the structure, in order to achieve biological function, appears to be a common property of proteins, as also evidenced for ubiquitin in a recent study [173].

4.0 CONCLUSION

In this chapter, we summarize the results found in the previous two chapters, discuss the limitation of our models, acknowledge the current challenges, and provide guidelines for future studies.

4.1 STRUCTURAL SYMMETRY

In Chapter 2, we have demonstrated that the global modes can be expressed as combinations of vector spherical harmonics, based on the molecular shape and symmetry. The irreducible representations of icosahedral group Y have been related to the spherical harmonics Y_{lm}. The unit representation A of the icosahedral group is contained in $l = 0, 6, 10, 12$, etc. So the deformation corresponding to vector spherical harmonic type $\mathbf{V}_0$ (icosahedral type A) is nondegenerate because it preserves icosahedral symmetry. The allowed mode has degeneracy $\Omega_0 = 1$ (see Eq. (2.16)). The $\mathbf{V}_0$ type defomation, a radial shrinking or swelling, would be expected to occur in response to strong internal pressure (such as the packaging of the genome in a bacteriophage) or external pressure (such as osmotic pressure).

Such icosahedrally symmetric modes have been shown to strongly correlate with the structural changes occuring during capsid maturation [36]. It is noteworthy that the eigenvalue of

the $\mathbf{V}_0$ type mode is substantially higher than the $\mathbf{V}_2 + \mathbf{W}_2$ type (squeezing mode), implying relative resistance against these naturally occurring forces. Because of the high symmetry it is easy to calculate the response of the shell to pressure differentials. For example, the increase in the outer radius of a spherical shell, subject to internal pressure p is [111]

$$u = \left(\frac{pR_1^3 R_2}{R_2^3 - R_1^3} \right) \left(\frac{3(1-\sigma)}{2E} \right), \tag{4.1}$$

Experimental examination of capsid dimensions in response to changes in pressure can thus provide information about material parameters.

4.2 MACROSCOPIC ELASTIC PROPERTIES

Recently, Young's moduli have been obtained from nanoindentation experiments. Bacteriophage $\phi 29$ has a value of $E = 1.8 \pm 0.2$ GPa [174], which is close to hard plastics and other proteins such as actin, tubulin [175] and lysozyme crystals [176, 177]. The cowpea chlorotic mottle virus is soft in comparison, with Young's moduli of 140 MPa and 190 MPa for the wild type and mutants respectively [125], which are similar in magnitude to soft plastics and Teflon [175]. For murine leukemia virus particles, the Young's moduli of immature and mature capsids are 0.23 GPa and 1.03 GPa [126], which are comparable to the moduli of the bacteriophage $\phi 29$ and microtubules (0.8 GPa) [178].

A longitudinal sound speed of $c_l = 1817$ m/s was measured by ultrasound in lysozyme single crystals [176]. Assuming Poisson's ratio $\sigma = 1/3$ yields a predicted transverse sound speed of $c_t = 915$ m/s. Brillouin scattering measurements [179] yielded sound velocities of

STMV capsids ranging from $c_l = 1920$ m/s (fully hydrated) to $c_l = 3350$ m/s (fully dehydrated). Assuming $\sigma = 1/3$, these sound speeds for STMV imply Young's moduli of 3.7 and 11.2 GPa, respectively.

We note that a generic value for the GNM force constant has been deduced by Phillips and coworkers [180], based on the comparison of predicted residue fluctuations with experimental B-factors. Accordingly, the force constant is given by $k_B T / \gamma = 0.87 \pm 0.46$ Å^2, or $\gamma = 0.48 \pm 0.35$ N/m. Using this value together with the ratio $E / \gamma = 26$ nm^{-1} obtained here for STMV (Table 3), the Young's modulus becomes $E = 12 \pm 9$ GPa, which is comparable to that inferred from Brillouin scattering data.

Note that the Poisson's ratio is unaffected by a scaling of γ, hence our fitting serves as a prediction of the value of σ. Experimentally, σ has been hard to obtain for biological molecules, and values around 0.3 have been assumed [174] by analogy to similar materials. Inspecting Table 3, we see predicted values in the range of 0.2-0.3.

4.3 COUPLING BETWEEN DYNAMICS AND PROTEIN ALLOSTERY[3]

Figure 18 displays the most cooperative mode of motion found by the ANM analysis of GroEL dynamics. The low frequency modes predicted by the ANM, or EN models in general, have been shown in numerous studies to carry information on functional motions. An early application of ANM to GroEL-GroES showed that the most cooperative motion of the complex is a global torsion where the two rings tend to rotate in opposite directions [32], as illustrated in the color-

[3] The content of the section 4.3 has been published on *Mol. BioSys*, 2008 [13].

coded ribbon diagrams in Figure 18. This motion involves the cooperative (all-or-none) conformational transition of all subunits between open and closed forms. A recent study by Thirumalai and coworkers also demonstrated how this single mode intrinsically accessible to the GroEL-$(ADP)_7$-GroES structure (R"T state) is insensitive to sequence variations, and dominates the transition between the R”T and TR” states [145].

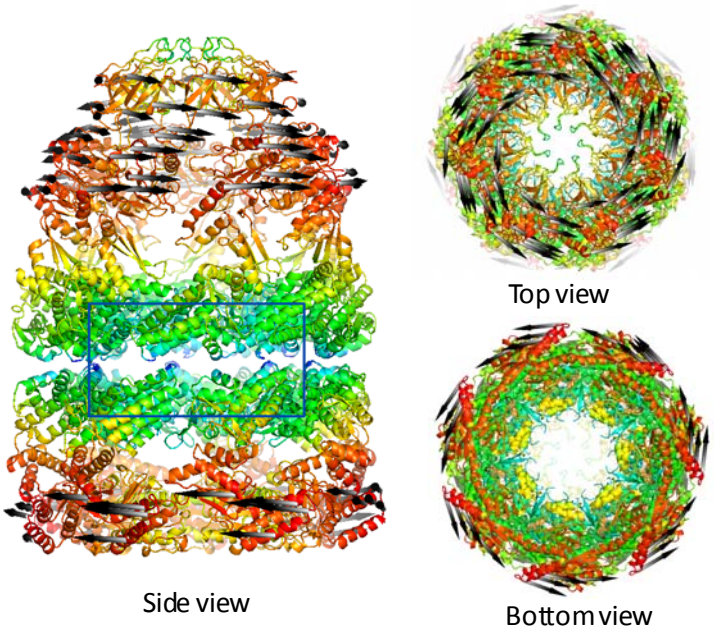

Figure 18. The first nonzero eigenmode of GroEL-GroES complex.

The slowest mode of GroEL-GroES complex is torsional motion around the cylindrical axis of symmetry. The residues are colored according to their mobilities, red indicating highly mobile and blue very rigid. The arrows indicate the counter-rotations of *cis* and *trans* rings. The boxed area is the region of interest, where E461 on the *cis* ring forms a salt bridge with residue R452 on the *trans* ring.

Two key interactions, A109-K105 (K-D interface) and R452-E461 (K-E interface), which mediate the inter-ring communication, have been identified by Markov propagation method [27]. E461 on the *cis* ring subunit forms a salt bridge with R452 on the *trans* ring which seems to be crucial for signal transmission (see Figure 19). The substitution of E461K has been reported by

Saibil and coworkers to impede the release of co-chaperonin GroES, and this long-range effect has been attributed to the redistribution of salt bridges at the interface between the two rings, induced by the counter-rotation of the two rings [151].

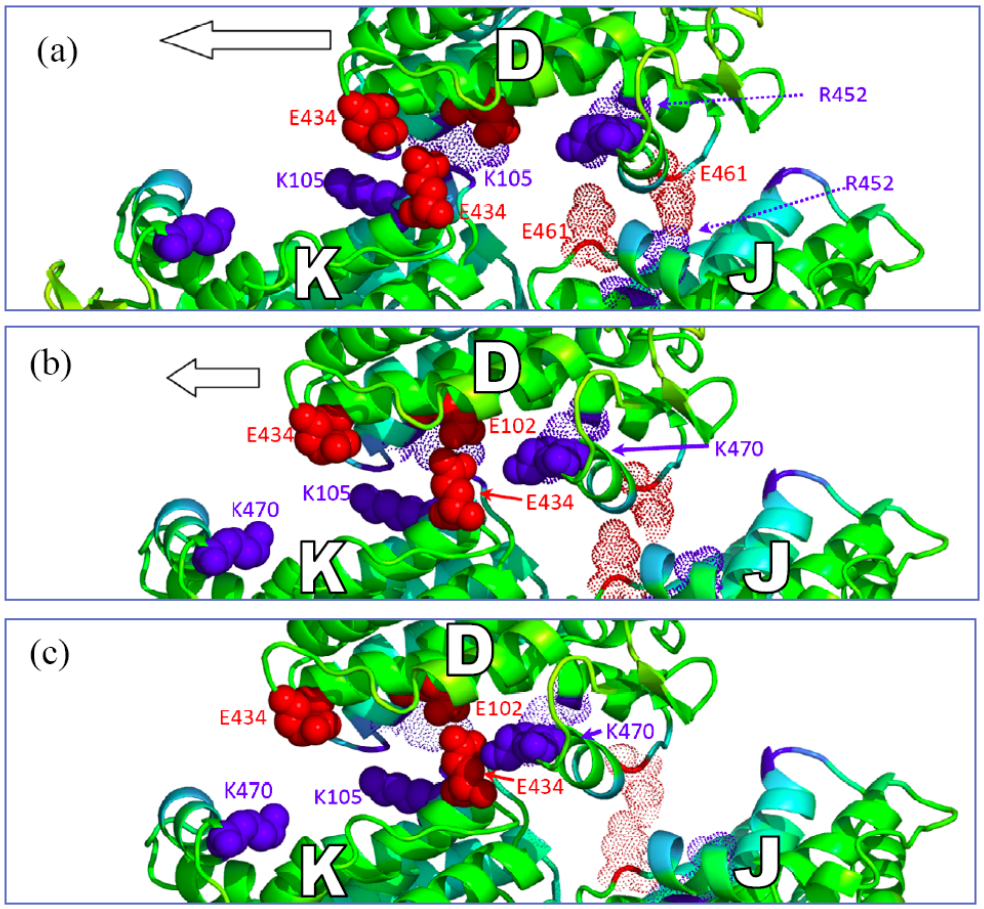

Figure 19. Redistribution of interactions at the interface of *cis* and *trans* rings, caused by the global motion of the chaperonin GroEL-GroES.

(A) Conformation prior to deformation (in the structure determined by Xu et al.[144]). We note that the subunits assume an out-of-register conformation: the subunit in the *trans* ring (e.g. D) is simultaneously interacting with two subunits in the *cis* ring (labeled K and J). In this conformation we note that there are four inter-ring hydrogen bonds: R452(J)– E461(D), R452(D)-E461(J), K105(D)-E434(K) and E434(D)-K105(K), from right to left. The acidic residues are colored red, and basic residues, blue. Dotted and filled coloring scheme of the atoms aims at facilitating the visualization of the redistribution of interactions. (**B** and **C**) Two successive conformations visited as the structure undergoes the global torsional motion shown in Figure 18. Note that E434(K) changes its partner from

K105(D) to K470 (D), consistent with the observations reported on the mutant E461K by Saibil and coworkers. At the same time E102 (D) approaches K105(K), and E434(D) approaches K470(K), suggesting that these oppositely charged residues potentially form salt bridges upon further deformation along this mode.

Interestingly, the lowest frequency mode predicted by the EN also shows a torsional rotation of the two GroEL rings (Figure 19). Of interest is to examine what happens at the interface during this motion. Figure 19 gives a closer view of the interactions at the interface before (Figure 19A) and after (Figure 19 panels B and C) the conformational motion driven by mode 1. The net effect of this global mode emerges as a redistribution of the interactions at the interface in agreement with those occurring in the E461K mutant. See Figure 19 caption for more details.

This observation suggests that the rearrangements observed in the E461K are simply those along the first global mode accessible to the complex even in the absence of mutation. We note that the global mode is in principle a symmetric oscillation between two conformers along opposite directions of the first principal axis (first nonzero mode). Therefore, according to the ANM, the molecule is equally able to undergo global deformations in either direction. It remains to be seen if the opposite direction torsion is stabilized under different conditions.

4.4 LIMITATIONS AND IMPROVEMENTS

Although the *a*ANM is an efficient method for revealing the highly probable pathways, which are beyond the range of exploration of other computational tools, the method has its own limitations. As we have discussed the method assumption in section 3.1.2. The modes predicted

by the ANM are those exclusively based on inter-residue contact topology. No specific interactions are taken into consideration. So, the routes predicted here are those selected assuming that mechanical effects purely based on geometry dominate the dynamics. However, specific interactions may become more important as more localized changes are simulated, and these are usually manifested by changes in side chain reorientations, which are beyond the range of *a*ANM calculations.

Proteins are moving on complicated energy landscape dictated by the Born-Oppenheimer approximation, which would involve solving the electronic Schrodinger equation. To make it feasible, empirical energy functions were developed, such as the ones used in CHARMM [181] and AMBER [182]. In the future works, the level of coarse-graining will be adjustable. Especially in studying the small to middle size proteins, and permitted by the computing resources, detailed interactions will be taken into consideration.

Due to the limits of applicability of the harmonic potential, it is arguable how far away from the original energy minima we can extend the quadratic approximation. Since the actual energy landscape is rugged. The system can find local minima everywhere, by reorienting its side-chains and adjusting local interactions. In my future work to modify the method, an energy minimization to relax the system after each iteration will be preformed. Alternatively, Miloshevsky and Jordan provided a solution to handle the harmonic approximation away from the energy minima [100, 102].

The *a*ANM does not contain ‘time’ explicitly. We note in particular that there are well-defined ways to define transition rates through reactive flux theory [183], optimize reaction coordinates and estimate reaction rate coefficients (see for example [184]), which are not addressed in the present study.

As discussed in section 3.3.2.3, the structural symmetry is kept during the transition. The symmetric modes can be taken a step further, and solely generated by molecular symmetry with group theory [35]. By combining these modes as we did in *a*ANM, the calculation can be simplified further, especially in studying the large size macromolecules with defined symmetry, e.g. the maturation of bacteriophage HK97 capsid [34].

Typically, in the protein-molecule docking problem, both the unbound and bound structures may have been determined by experiments. To explore the binding mechanism or refine the binding site, solely using the ending structures is not enough. With *a*ANM, we can readily generate a series of intermediate structures which can be fed as input to docking simulation tools.

APPENDIX A

BOUNDARY CONDITIONS

Spherical symmetry ensures that oscillation frequencies depend on the angular momentum index l but are independent of the azimuthal quantum number m. Since m ranges between $-l$ and $+l$, a mode of angular momentum index l has degeneracy $2l+1$. The value $m=0$ yields modes of specially simple form because $\mathbf{u}_0$ becomes independent of the azimuthal angle ϕ. Specifically,

$$\begin{aligned}\mathbf{u}_0(\boldsymbol{r}) = c_0\left(qj'(qr)Y_l(\theta)\hat{\boldsymbol{r}} + \frac{j(qr)}{r}\frac{\partial Y_l(\theta)}{\partial\theta}\hat{\boldsymbol{\theta}}\right) + c_1\left(j(kr)\frac{\partial Y_l(\theta)}{\partial\theta}\hat{\boldsymbol{\phi}}\right) \\ + c_2\left(-l(l+1)\frac{j(kr)}{r}Y_l(\theta)\hat{\boldsymbol{r}} - \left[\frac{j(kr)}{r} + kj'(kr)\right]\frac{\partial Y_l(\theta)}{\partial\theta}\hat{\boldsymbol{\theta}}\right)\end{aligned}, \tag{A.1}$$

where the functions $Y_l(\theta) \equiv Y_{l0}(\theta,\phi)$ do not depend on ϕ, and c_0, c_1 and c_2 are as-yet undetermined constants. The coefficient c_1 multiplies the $\hat{\mathbf{f}}$ terms, tangent to the surface of the sphere, corresponding to *torsional* modes. The other two coefficients, c_0 and c_2 multiply the $\hat{\mathbf{r}}$ and $\hat{\mathbf{q}}$ terms, and correspond to *spheroidal* modes.

Substituting the above general solution into the boundary condition Eq. (2.14), we obtain

$$\begin{bmatrix} A_{11} & 0 & A_{13} \\ 0 & A_{22} & 0 \\ A_{31} & 0 & A_{33} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} \Bigg\|_{r=R} = 0 \tag{A.2}$$

where

$$\begin{aligned} A_{11} = &\quad (1-\sigma)q^2 j''(qr) + 2\sigma \frac{qj'(qr)}{r} - \sigma l(l+1)\frac{j(qr)}{r^2} \\ A_{13} = &\quad (2\sigma - 1)l(l+1)\left(\frac{kj'(kr)}{r} - \frac{j(kr)}{r^2}\right) \\ A_{22} = &\quad kj'(kr) - \frac{j(kr)}{r} \\ A_{31} = &\quad 2\left(\frac{qj'(qr)}{r} - \frac{j(qr)}{r^2}\right) \\ A_{33} = &\quad \left(2 - l(l+1)\right)\frac{j(kr)}{r^2} - k^2 j''(kr) \end{aligned} \quad . \tag{A.3}$$

Note that in Eq. (A.2) the coefficient matrix separates into two sub-matrices. One submatrix multiplies the coefficient c_1, the other multiplies c_0 and c_2. Nonzero solutions to Eq. (A.2) exist only if the determinant of the matrix vanishes. Since the matrix divides into two sub-matrices, the solvability condition becomes

$$A_{22}\big|_{r=R} = 0 \tag{A.4}$$

and

$$\det \begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{vmatrix}_{r=R} = 0 \,. \tag{A.5}$$

By solving the above equations, for each angular momentum index l, the transverse wavenumber k can be extracted, yielding the vibrational frequency $\omega = c_t k$ and the "eigenvalue" $\lambda^{(cont)} = \omega^2$.

APPENDIX B

STEEPEST DESCENT PATH

The steepest descent path (SDP) between the conformations $\boldsymbol{R}_A^{(0)}$ and $\boldsymbol{R}_B^{(0)}$ is calculated by numerical minimization of the integral

$$S = \int_{\boldsymbol{R}_A^{(0)}}^{\boldsymbol{R}_B^{(0)}} \left|\nabla U\right| dl. \tag{B.1}$$

The product $\left|\nabla U\right| dl \geq \left|\nabla U \mathbf{dl}\right|$, therefore $\int_{\boldsymbol{R}_A^{(0)}}^{\boldsymbol{R}_B^{(0)}} \left|\nabla U\right| dl \geq \int_{\boldsymbol{R}_A^{(0)}}^{\boldsymbol{R}_B^{(0)}} \left|\nabla U \mathbf{dl}\right| = \sum_i \left|\Delta U_i\right|$, where ΔU_i is potential energy difference in the i^{th} monotonic part of the trajectory. The sum $\sum_i \left|\Delta U_i\right|$ is minimal if the molecule undergoes a transition from $\boldsymbol{R}_A^{(0)}$ to $\boldsymbol{R}_B^{(0)}$ through the lowest energy barrier; moreover, the terms in the last inequality are equal if and only if the trajectory is parallel to the potential gradient. Therefore SDP is the unique minimum of the boundary value integral in Eq. (B.1). To find a path that minimizes S (i.e., the SDP), we search for the global minimum of a discrete approximation

$$S'\left(\boldsymbol{R}_2, \ldots, \boldsymbol{R}_{n-1}\right) = \sum_{i=1}^{n-1} \left|\nabla U_{|\boldsymbol{R}_i}\right| \left|\boldsymbol{R}_{i+1} - \boldsymbol{R}_i\right|, \tag{B.2}$$

subject to the constraint that successive $\boldsymbol{R}_i$'s are equally spaced along the trajectory. The advantage of this algorithm is that it remains stable and calculates qualitatively reasonable trajectories even when the distances between $\boldsymbol{R}_i$'s are large. The minimization of S' is accomplished by simulated annealing. For further details see [59] and [62].

APPENDIX C

RELATIONSHIP BETWEEN CUMULATIVE CORRELATION COSINE AND DEVIATION ANGLE

The instantaneous distance vector $\boldsymbol{d}^{(k)}$ can be expressed as:

$$\boldsymbol{d}^{(k)} = \boldsymbol{R}_B^{(k)} - \boldsymbol{R}_A^{(k)} = \sum_{i=1}^{3N-6} \alpha_i \boldsymbol{u}_{iA}^{(k)} \tag{C.1}$$

The instantaneous deformation vector $\boldsymbol{v}_A^{(k)}$ can, in turn, be written as the summation of weighted eigenvectors (see Eq. (3.4)) as

$$\boldsymbol{v}_A^{(k)} = \boldsymbol{R}_A^{(k)} - \boldsymbol{R}_A^{(k-1)} = s_A^{(k)} \sum_{j=1}^{m_A^{(k)}} \alpha_{jA}^{(k)} \, \boldsymbol{u}_{jA}^{(k)} \tag{C.2}$$

We will conveniently omit the subscript A and superscript k in the following, as the derivation holds for all steps (k) and starting conformation (A or B). The dot product of the vectors $\boldsymbol{d}$ and $\boldsymbol{v}$ can be written in terms of the summations over nonzero modes as

$$\boldsymbol{d} \cdot \boldsymbol{v} = \left(\sum_{i=1}^{3N-6} \alpha_i \boldsymbol{u}_i \right) \cdot \left(s \sum_{j=1}^{m} \alpha_j \boldsymbol{u}_j \right) = s \sum_{j=1}^{m} \alpha_i^2 \left| \boldsymbol{u}_i \right|^2 \tag{C.3}$$

The equality follows from the orthogonality of the eigenvectors, i.e., $\boldsymbol{u}_i \cdot \boldsymbol{u}_j = \delta_{ij}$. Also, since the eigenvectors are normalized, $\left| \boldsymbol{u}_i \right| = 1$, we obtain

$$\boldsymbol{d} \cdot \boldsymbol{v} = s \sum_{i=1}^{m} \alpha_i^2 \tag{C.4}$$

Using the definitions $\alpha_i = |\boldsymbol{d}| cos(\boldsymbol{d}, \boldsymbol{u}_i)$ and $F(m) = \sum_{i=1}^{m} \left(cos(\boldsymbol{d}, \boldsymbol{u}_i) \right)^2$ in Eq.(C.4), we obtain

$$\boldsymbol{d} \cdot \boldsymbol{v} = s \sum_{i=1}^{m} \left(|\boldsymbol{d}| cos(\boldsymbol{d}, \boldsymbol{u}_i) \right)^2 = s |\boldsymbol{d}|^2 F(m) \tag{C.5}$$

On the other hand,

$$|\boldsymbol{v}|^2 = \left(s \sum_{i=1}^{m} \alpha_i \boldsymbol{u}_i \right) \cdot \left(s \sum_{i=1}^{m} \alpha_i \boldsymbol{u}_i \right) = s^2 \sum_{i=1}^{m} \alpha_i^2 |\boldsymbol{u}_i|^2 = s^2 \sum_{i=1}^{m} \alpha_i^2 \tag{C.6}$$

Here we have utilized both orthogonality and unit magnitude properties of the eigenvectors, $\boldsymbol{u}$.

Using the definitions of α_i and $F(m)$, this equality becomes

$$|\boldsymbol{v}|^2 = s^2 \sum_{i=1}^{m} \left(|\boldsymbol{d}| cos(\boldsymbol{d}, \boldsymbol{u}_i) \right)^2 = s^2 |\boldsymbol{d}|^2 F(m) \tag{C.7}$$

which implies $|\boldsymbol{v}| = s |\boldsymbol{d}| \left[F(m) \right]^{1/2}$. Combining this result with Eq. (C.5) leads to Eq. (3.6).

BIBLIOGRAPHY

1. Voet DJ, Voet JG, & Pratt CW (2008) *Fundamentals of Biochemistry: Life at the Molecular Level* (Wiley, Hoboken, NJ).
2. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, & Phillips DC (1958) A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181**: 662-666.
3. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, & Bourne PE (2000) The Protein Data Bank. *Nucl. Acids Res.* **28**: 235-242.
4. Kurland CG (1992) Translational Accuracy and the Fitness of Bacteria. *Annu Rev Genet* **26**: 29-50.
5. Stossel T (1989) From Signal to Pseudopod. How Cells Control Cytoplasmic Actin Assembly. *J Biol Chem* **264**: 18261-18264.
6. Goodsell DS & Olson AJ (2000) Structural Symmetry and Protein Function. *Annu Rev Biophys Biomol Struct* **29**: 105-153.
7. Cornish-Bowden AJ & Koshland DE, Jr. (1971) The Quaternary Structure of Proteins Composed of Identical Subunits. *J Biol Chem* **246**: 3092-3102.
8. Wolynes PG (1996) Symmetry and the Energy Landscapes of Biomolecules. *Proc Natl Acad Sci U S A* **93**: 14249-14255.
9. Changeux JP & Edelstein SJ (2005) Allosteric Mechanisms of Signal Transduction. *Science* **308**: 1424-1428.
10. Monod J, Wyman J, & Changeux JP (1965) On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol* **12**: 88-8118.
11. Xu C, Tobi D, & Bahar I (2003) Allosteric Changes in Protein Structure Computed by a Simple Mechanical Model: Hemoglobin T<-->R2 Transition. *J Mol Biol* **333**: 153-168.
12. Koshland DE, Némethy G, & Filmer D (1966) Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **5**: 365-385.
13. Chennubhotla C, Yang Z, & Bahar I (2008) Coupling between Global Dynamics and Signal Transduction Pathways: A Mechanism of Allostery for Chaperonin GroEL. *Mol BioSyst* **4**: 287-292.
14. Bahar I & Rader AJ (2005) Coarse-Grained Normal Mode Analysis in Structural Biology. *Curr Opin Struct Biol* **15**: 586-592.
15. Cui Q & Bahar I (2006) *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems* (Chapman & Hall/CRC, Boca Raton, FL).
16. Yu H, Ma L, Yang Y, & Cui Q (2007) Mechanochemical Coupling in the Myosin Motor Domain. II. Analysis of Critical Residues. *PLoS Comput Biol* **3**: e23.
17. Chennubhotla C & Bahar I (2007) Signal Propagation in Proteins and Relation to Equilibrium Fluctuations. *PLoS Comput Biol* **3**: e172.

18. Bahar I, Chennubhotla C, & Tobi D (2007) Intrinsic Dynamics of Enzymes in the Unbound State and Relation to Allosteric Regulation. *Curr Opin Struct Biol* **17**: 633-640.
19. Bahar I, Atilgan AR, & Erman B (1997) Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential. *Fold. Des.* **2**: 173-181.
20. Doruker P, Atilgan AR, & Bahar I (2000) Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to α-Amylase Inhibitor. *Proteins* **40**: 512-524.
21. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, & Bahar I (2001) Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys J* **80**: 505-515.
22. Durand P, Trinquier G, & Sanejouand YH (1994) New Approach for Determining Low-Frequency Normal-Modes in Macromolecules. *Biopolymers* **34**: 759-771.
23. Tama F, Gadea FX, Marques O, & Sanejouand YH (2000) Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules. *Proteins-Structure Function and Genetics* **41**: 1-7.
24. Li GH & Cui Q (2002) A Coarse-Grained Normal Mode Approach for Macromolecules: An Efficient Implementation and Application to Ca^{2+}-ATPase. *Biophys J* **83**: 2457-2474.
25. Li G & Cui Q (2004) Analysis of Functional Motions in Brownian Molecular Machines with an Efficient Block Normal Mode Approach: Myosin-II and Ca^{2+} -ATPase. *Biophys J* **86**: 743-763.
26. Doruker P, Jernigan RL, & Bahar I (2002) Dynamics of Large Proteins through Hierarchical Levels of Coarse-Grained Structures. *J Comput Chem* **23**: 119-127.
27. Chennubhotla C & Bahar I (2006) Markov Propagation of Allosteric Effects in Biomolecular Systems: Application to GroEL-GroES. *Mol Syst Biol* **2**: 1-13.
28. Kurkcuoglu O, Jernigan RL, & Doruker P (2005) Collective Dynamics of Large Proteins from Mixed Coarse-Grained Elastic Network Model. *QSAR & Combinatorial Science* **24**: 443-448.
29. Micheletti C, Carloni P, & Maritan A (2004) Accurate and Efficient Description of Protein Vibrational Dynamics: Comparing Molecular Dynamics and Gaussian Models. *Proteins: Structure, Function, and Bioinformatics* **55**: 635-645.
30. Tama F, Valle M, Frank J, & Brooks CL, 3rd (2003) Dynamic Reorganization of the Functionally Active Ribosome Explored by Normal Mode Analysis and Cryo-Electron Microscopy. *Proc Natl Acad Sci U S A* **100**: 9319-9323.
31. Wang Y, Rader AJ, Bahar I, & Jernigan RL (2004) Global Ribosome Motions Revealed with Elastic Network Model. *J Struct Biol* **147**: 302-314.
32. Keskin O, Bahar I, Flatow D, Covell DG, & Jernigan RL (2002) Molecular Mechanisms of Chaperonin GroEL-GroES Function. *Biochemistry* **41**: 491-501.
33. Ma J, Sigler PB, Xu Z, & Karplus M (2000) A Dynamic Model for the Allosteric Mechanism of GroEL. *J Mol Biol* **302**: 303-313.
34. Rader AJ, Vlad DH, & Bahar I (2005) Maturation Dynamics of Bacteriophage HK97 Capsid. *Structure* **13**: 413-421.
35. van Vlijmen HW & Karplus M (2005) Normal Mode Calculations of Icosahedral Viruses with Full Dihedral Flexibility by Use of Molecular Symmetry. *J Mol Biol* **350**: 528-542.
36. Tama F & Brooks CLI (2005) Diversity and Identity of Mechanical Properties of Icosahedral Viral Capsids Studied with Elastic Network Normal Mode Analysis *J Mol Biol* **345**: 299-314.

37. Ming DM, Kong YF, Wakil SJ, Brink J, & Ma JP (2002) Domain Movements in Human Fatty Acid Synthase by Quantized Elastic Deformational Model. *Proc Natl Acad Sci U S A* **99**: 7895-7899.
38. Ming D, Kong YF, Lambert MA, Huang Z, & Ma JP (2002) How to Describe Protein Motion without Amino Acid Sequence and Atomic Coordinates. *Proc Natl Acad Sci U S A* **99**: 8620-8625.
39. Delarue M & Dumas P (2004) On the Use of Low-Frequency Normal Modes to Enforce Collective Movements in Refining Macromolecular Structural Models. *Proc Natl Acad Sci U S A* **101**: 6957-6962.
40. Hinsen K, Reuter N, Navaza J, Stokes DL, & Lacapere JJ (2005) Normal Mode-Based Fitting of Atomic Structure into Electron Density Maps: Application to Sarcoplasmic Reticulum Ca-ATPase. *Biophys J* **88**: 818-827.
41. Zhang X, Meining W, Cushman M, Haase I, Fischer M, Bacher A, & Ladenstein R (2003) A Structure-Based Model of the Reaction Catalyzed by Lumazine Synthase from Aquifex Aeolicus. *J Mol Biol* **328**: 167-182.
42. Tama F & Brooks CL (2006) Symmetry, Form, and Shape: Guiding Principles for Robustness in Macromolecular Machines. *Annu Rev Biophys Biomol Struct* **35**: 115-133.
43. Ma J (2005) Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* **13**: 373-380.
44. Nicolay S & Sanejouand YH (2006) Functional Modes of Proteins Are among the Most Robust. *Phys Rev Lett* **96**: 078104-078104.
45. Arnoldi WE (1951) The Principle of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem. *Quarterly of Applied Mathematics* **9**: 17–29.
46. Lehoucq RBS, D.C. & Yang C (1998) *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods* (SIAM).
47. Vinson VJ (2009) Proteins in Motion. Introduction. *Science* **324**: 197.
48. Smock RG & Gierasch LM (2009) Sending Signals Dynamically. *Science* **324**: 198-203.
49. Tokuriki N & Tawfik DS (2009) Protein Dynamism and Evolvability. *Science* **324**: 203-207.
50. Engler AJ, Humbert PO, Wehrle-Haller B, & Weaver VM (2009) Multiscale Modeling of Form and Function. *Science* **324**: 208-212.
51. Bolhuis PG, Chandler D, Dellago C, & Geissler PL (2002) Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **53**: 291-318.
52. van der Vaart A (2006) Simulation of Conformational Transitions. *Theor Chem Acc* **116**: 183-193.
53. Elber R & Karplus M (1987) A Method for Determining Reaction Paths in Large Molecules: Application to Myoglobin. *Chem Phys Lett* **139**: 375-380.
54. Czerminski R & Elber R (1990) Reaction Path Study of Conformational Transitions in Flexible Systems: Applications to Peptides. *J. Chem. Phys.* **92**: 5580-5601.
55. Olender R & Elber R (1996) Calculation of Classical Trajectories with a Very Large Time Step: Formalism and Numerical Examples. *J. Chem. Phys.* **105**: 9299-9315.
56. Huo S & Straub JE (1997) The MaxFlux Algorithm for Calculating Variationally Optimized Reaction Paths for Conformational Transitions in Many Body Systems at Finite Temperature. *J. Chem. Phys.* **107**: 5000-5006.

57. Yang H, Wu H, Li D, Han L, & Huo S (2007) Temperature-Dependent Probabilistic Roadmap Algorithm for Calculating Variationally Optimized Conformational Transition Pathways. *J Chem Theory Comp* **3**: 17-25.
58. Jónsson H, Mills G, & Jacobsen KW (1998) Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions in *Classical and Quantum Dynamics in Condensed Phase Simulations*, eds. B. J. Berne GC & Coker DF (World Scientific, Singapore), pp. 385–404.
59. Elber R & Shalloway D (2000) Temperature Dependent Reaction Coordinates. *J. Chem. Phys.* **112**: 5539-5545.
60. Elber R, Ghosh A, & Cárdenas A (2002) Long Time Dynamics of Complex Systems. *Acc Chem Res* **35**: 396-403.
61. Elber R (2005) Long-Timescale Simulation Methods. *Curr Opin Struct Biol* **15**: 151-156.
62. Májek P, Elber R, & Weinstein H (2008) Pathways of Conformational Transitions in Proteins in *Coarse-Graining of Condensed Phase and Biomolecular Systems*, ed. Voth GA (CRC Press, Boca Raton, FL).
63. Vonrhein C, Schlauderer GJ, & Schulz GE (1995) Movie of the Structural Changes During a Catalytic Cycle of Nucleoside Monophosphate Kinases. *Structure* **3**: 483-490.
64. Echols N, Milburn D, & Gerstein M (2003) MolMovDB: Analysis and Visualization of Conformational Change and Structural Flexibility. *Nucleic Acids Res* **31**: 478-482.
65. Schlitter J, Engels M, Krueger P, Jacoby E, & Wollmer A (1993) Targeted Molecular Dynamics Simulation of Conformational Change: Application to the T-R Transition in Insulin. *Mol Simulat* **10**: 291-308.
66. Ma J & Karplus M (1997) Molecular Switch in Signal Transduction: Reaction Paths of the Conformational Changes in *Ras* p21. *Proc Natl Acad Sci U S A* **94**: 11905-11910.
67. Ma J, Flynn TC, Cui Q, Leslie AGW, Walker JE, & Karplus M (2002) A Dynamic Analysis of the Rotation Mechanism for Conformational Change in F_1-ATPase. *Structure* **10**: 921-931.
68. Flynn TC, Swint-Kruse L, Kong Y, Booth C, Matthews KS, & Ma J (2003) Allosteric Transition Pathways in the Lactose Repressor Protein Core Domains: Asymmetric Motions in a Homodimer. *Protein Sci* **12**: 2523-2541.
69. Zhang J, Li C, Chen K, Zhu W, Shen X, & Jiang H (2006) Conformational Transition Pathway in the Allosteric Process of Human Glucokinase. *Proc Natl Acad Sci U S A* **103**: 13368-13373.
70. Grubmuller H, Heymann B, & Tavan P (1996) Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **271**: 997-999.
71. Gu Y, Shrivastava IH, Amara SG, & Bahar I (2009) Molecular Simulations Elucidate the Substrate Translocation Pathway in a Glutamate Transporter. *Proceedings of the National Academy of Sciences* **106**: 2589-2594.
72. Pratt LR (1986) A Statistical Method for Identifying Transition States in High Dimensional Problems. *The Journal of Chemical Physics* **85**: 5045-5048.
73. Dellago C, Bolhuis PG, & Chandler D (1998) Efficient Transition Path Sampling: Application to Lennard-Jones Cluster Rearrangements. *J. Chem. Phys.* **108**: 9236-9245.
74. Dellago C, Bolhuis PG, Csajka FS, & Chandler D (1998) Transition Path Sampling and the Calculation of Rate Constants. *J. Chem. Phys.* **108**: 1964-1977.
75. Paniconi M & Zimmer MF (1999) Statistical Features of Large Fluctuations in Stochastic Systems. *Phys Rev E* **59**: 1563.

76. Oettinger HC (1994) Variance Reduced Brownian Dynamics Simulations. *Macromolecules* **27**: 3415-3423.
77. Woolf TB (1998) Path Corrected Functionals of Stochastic Trajectories: Towards Relative Free Energy and Reaction Coordinate Calculations. *Chem Phys Lett* **289**: 433-441.
78. Zuckerman DM & Woolf TB (1999) Dynamic Reaction Paths and Rates through Importance-Sampled Stochastic Dynamics. *J. Chem. Phys.* **111**: 9475-9484.
79. Zuckerman DM & Woolf TB (2001) Efficient Dynamic Importance Sampling of Rare Events in One Dimension. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **63**: 016702.
80. Huber GA & Kim S (1996) Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophys J* **70**: 97-110.
81. Rojnuckarin A, Kim S, & Subramaniam S (1998) Brownian Dynamics Simulations of Protein Folding: Access to Milliseconds Time Scale and Beyond. *Proc Natl Acad Sci U S A* **95**: 4288-4292.
82. Rojnuckarin A, Livesay DR, & Subramaniam S (2000) Bimolecular Reaction Simulation Using Weighted Ensemble Brownian Dynamics and the University of Houston Brownian Dynamics Program. *Biophys J* **79**: 686-693.
83. Fisher EW, Rojnuckarin A, & Kim S (2001) Kinetic Effects of Mutations of Charged Residues on the Surface of a Dimeric Hemoglobin: Insights from Brownian Dynamics Simulations. *J Mol Stuct: THEOCHEM* **549**: 47-54.
84. Fisher EW, Rojnuckarin A, & Kim S (2002) Exhaustive Enumeration of the Effects of Point Charge Mutations on the Electrostatically Driven Association of Hemoglobin Subunits, Using Weighted-Ensemble Brownian Dynamics Simulations. *Structural Chemistry* **13**: 193-202.
85. Zhang BW, Jasnow D, & Zuckerman DM (2007) Efficient and Verified Simulation of a Path Ensemble for Conformational Change in a United-Residue Model of Calmodulin. *Proceedings of the National Academy of Sciences* **104**: 18043-18048.
86. Tama F & Sanejouand YH (2001) Conformational Change of Proteins Arising from Normal Mode Calculations. *Protein Eng* **14**: 1-6.
87. Kim MK, Jernigan RL, & Chirikjian GS (2002) Efficient Generation of Feasible Pathways for Protein Conformational Transitions. *Biophys J* **83**: 1620-1630.
88. Kim MK, Chirikjian GS, & Jernigan RL (2002) Elastic Models of Conformational Transitions in Macromolecules. *J Mol Graph Model* **21**: 151-160.
89. Kim MK, Jernigan RL, & Chirikjian GS (2005) Rigid-Cluster Models of Conformational Transitions in Macromolecular Machines and Assemblies. *Biophys J* **89**: 43-55.
90. Zheng W & Brooks BR (2005) Normal-Modes-Based Prediction of Protein Conformational Changes Guided by Distance Constraints. *Biophys J* **88**: 3109-3117.
91. Zheng W & Brooks BR (2006) Modeling Protein Conformational Changes by Iterative Fitting of Distance Constraints Using Reoriented Normal Modes. *Biophys J* **90**: 4327-4336.
92. Miyashita O, Onuchic JN, & Wolynes PG (2003) Nonlinear Elasticity, Proteinquakes, and the Energy Landscapes of Functional Transitions in Proteins. *Proc Natl Acad Sci U S A* **100**: 12570-12575.
93. Whitford PC, Miyashita O, Levy Y, & Onuchic JN (2007) Conformational Transitions of Adenylate Kinase: Switching by Cracking. *J Mol Biol* **366**: 1661--1671.

94. Maragakis P & Karplus M (2005) Large Amplitude Conformational Change in Proteins Explored with a Plastic Network Model: Adenylate Kinase *J Mol Biol* **352**: 807-822
95. Best RB, Chen Y-G, & Hummer G (2005) Slow Protein Conformational Dynamics from Multiple Experimental Structures: The Helix/Sheet Transition of Arc Repressor. *Structure* **13**: 1755-1763.
96. Chu J-W & Voth GA (2007) Coarse-Grained Free Energy Functions for Studying Protein Conformational Changes: A Double-Well Network Model. *Biophys J* **93**: 3860-3871.
97. Isin B, Schulten K, Tajkhorshid E, & Bahar I (2008) Mechanism of Signal Propagation Upon Retinal Isomerization: Insights from Molecular Dynamics Simulations of Rhodopsin Restrained by Normal Modes. *Biophys J* **95**: 789-803.
98. Franklin J, Koehl P, Doniach S, & Delarue M (2007) MinActionPath: Maximum Likelihood Trajectory for Large-Scale Structural Transitions in a Coarse-Grained Locally Harmonic Energy Landscape. *Nucl. Acids Res.* **35**: W477-482.
99. Onsager L & Machlup S (1953) Fluctuations and Irreversible Processes. *Physical Review* **91**: 1505.
100. Miloshevsky GV & Jordan PC (2006) The Open State Gating Mechanism of Gramicidin a Requires Relative Opposed Monomer Rotation and Simultaneous Lateral Displacement. *Structure* **14**: 1241-1249.
101. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, & Teller E (1953) Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21**: 1087-1092.
102. Miloshevsky GV & Jordan PC (2007) Open-State Conformation of the KcsA K^+ Channel: Monte Carlo Normal Mode Following Simulations. *Structure* **15**: 1654-1662.
103. Yang Z, Májek P, & Bahar I (2009) Allosteric Transitions of Supramolecular Systems Explored by Network Models: Application to Chaperonin GroEL. *PLoS Comput Biol* **5**: e1000360.
104. Brooks B & Karplus M (1985) Normal Modes for Specific Motions of Macromolecules: Application to the Hinge-Bending Mode of Lysozyme. *Proc Natl Acad Sci U S A* **82**: 4995-4999.
105. Levitt M, Sander C, & Stern PS (1985) Protein Normal-Mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme. *J Mol Biol* **181**: 423-447.
106. Go N, Noguti T, & Nishikawa T (1983) Dynamics of a Small Globular Protein in Terms of Low-Frequency Vibrational Modes. *Proc Natl Acad Sci U S A* **80**: 3696-3700.
107. Yang Z, Bahar I, & Widom M (2009) Vibrational Dynamics of Icosahedrally Symmetric Biomolecular Assemblies Compared with Predictions Based on Continuum Elasticity. *Biophys J* **96**: 4438-4448.
108. Mackay AL (1962) A Dense Non-Crystallographic Packing of Equal Spheres. . *Acta Crystallogr.* **15**: 916-918.
109. Larson SB, Day J, Greenwood A, & McPherson A (1998) Refined Structure of Satellite Tobacco Mosaic Virus at 1.8 Angstrom Resolution. *J Mol Biol* **277**: 37-59.
110. Larson SB, Lucas RW, & McPherson A (2005) Crystallographic Structure of the T=1 Particle of Brome Mosaic Virus. *J Mol Biol* **346**: 815-831.
111. Landau LD & Lifshitz EM (1986) *Theory of Elasticity* (Butterworth-Heinemann).
112. Lobo JA (1995) What Can We Learn About Gravitational-Wave Physics with an Elastic Spherical Antenna. *Physical Review D* **52**: 591-604.

113. Coccia E, Fafone V, Frossati G, Lobo JA, & Ortega JA (1998) Hollow Sphere as a Detector of Gravitational Radiation. *Physical Review D* **57**: 2051-2060.
114. Lamb H (1881) On the Vibrations of an Elastic Sphere. *Proceedings of London Mathematical Society* **13**: 189-212.
115. Lamb H (1882) On the Vibrations of a Spherical Shell. *Proceedings of London Mathematical Society* **14**: 50-56.
116. Hill EL (1954) The Theory of Vector Spherical Harmonics. *American Journal of Physics* **22**: 211-214.
117. Steinhardt PJ, Nelson DR, & Ronchetti M (1983) Bond-Orientational Order in Liquids and Glasses. *Physical Review B* **28**: 784.
118. Yin ZY, Zheng YL, & Doerschuk PC (2001) An Ab Initio Algorithm for Low-Resolution 3-D Reconstructions from Cryoelectron Microscopy Images. *J Struct Biol* **133**: 132-142.
119. Tinkham M (1964) *Group Theory and Quantum Mechanics* (McGraw Hill).
120. Widom M, Lidmar J, & Nelson DR (2007) Soft Modes near the Buckling Transition of Icosahedral Shells. *Phys Rev E* **76**: 031911.
121. Caspar DL & A. K (1962) Physical Principles in the Construction of Regular Viruses. *Cold Spring Harbor Symp Quant Biol* **27**: 1-24.
122. Zandi R, Reguera D, Bruinsma RF, Gelbart WM, & Rudnick J (2004) Origin of Icosahedral Symmetry in Viruses. *Proc Natl Acad Sci U S A* **101**: 15556-15560.
123. Lorman VL & Rochal SB (2007) Density-Wave Theory of the Capsid Structure of Small Icosahedral Viruses. *Phys Rev Lett* **98**: 185502.
124. Lorman VL & Rochal SB (2008) Landau Theory of Crystallization and the Capsid Structures of Small Icosahedral Viruses. *Physical Review B* **77**: 224109.
125. Michel JP, Ivanovska IL, Gibbons MM, Klug WS, Knobler CM, Wuite GJ, & Schmidt CF (2006) Nanoindentation Studies of Full and Empty Viral Capsids and the Effects of Capsid Protein Mutations on Elasticity and Strength. *Proc Natl Acad Sci U S A* **103**: 6184-6189.
126. Kol N, Gladnikoff M, Barlam D, Shneck RZ, Rein A, & Rousso I (2006) Mechanical Properties of Murine Leukemia Virus Particles: Effect of Maturation. *Biophys J* **91**: 767-774.
127. Varma CM, Zaanen J, & Raghavachari K (1991) Superconductivity in the Fullerenes. *Science* **254**: 989-992.
128. Jishi RA & Dresselhaus MS (1992) Electron-Phonon Coupling Strength and Implications for Superconductivity in Alkali-Metal-Doped Fullerenes. *Phys Rev B Condens Matter* **45**: 2597-2600.
129. Eyal E & Bahar I (2008) Toward a Molecular Understanding of the Anisotropic Response of Proteins to External Forces: Insights from Elastic Network Models. *Biophys J* **94**: 3424-3435.
130. Dykeman EC & Sankey OF (2008) Low Frequency Mechanical Modes of Viral Capsids: An Atomistic Approach. *Phys Rev Lett* **100**: 028101.
131. Dykeman EC, Sankey OF, & Tsen KT (2007) Raman Intensity and Spectra Predictions for Cylindrical Viruses. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **76**: 011906.
132. Arkhipov A, Freddolino PL, & Schulten K (2006) Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling. *Structure* **14**: 1767-1777.

133. Freddolino PL, Arkhipov AS, Larson SB, McPherson A, & Schulten K (2006) Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. *Structure* **14**: 437-449.
134. Gibbons MM & Klug WS (2008) Influence of Nonuniform Geometry on Nanoindentation of Viral Capsids. *Biophys J* **95**: 3640-3649.
135. Gibbons MM & Klug WS (2007) Nonlinear Finite-Element Analysis of Nanoindentation of Viral Capsids. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **75**: 031901.
136. Guerin T & Bruinsma R (2007) Theory of Conformational Transitions of Viral Shells. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **76**: 061911.
137. Zandi R & Reguera D (2005) Mechanical Properties of Viral Capsids. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **72**: 021917.
138. Sigler PB, Xu ZH, Rye HS, Burston SG, Fenton WA, & Horwich AL (1998) Structure and Function in GroEL-Mediated Protein Folding. *Annu Rev Biochem* **67**: 581-608.
139. Saibil HR & Ranson NA (2002) The Chaperonin Folding Machine. *Trends Biochem Sci* **27**: 627-632.
140. Thirumalai D & Lorimer GH (2001) Chaperonin-Mediated Protein Folding. *Annu Rev Biophys Biomol Struct* **30**: 245-269.
141. Horovitz A & Willison KR (2005) Allosteric Regulation of Chaperonins. *Curr Opin Struct Biol* **15**: 646-651.
142. Ranson NA, Farr GW, Roseman AM, Gowen B, Fenton WA, Horwich AL, & Saibil HR (2001) ATP-Bound States of GroEL Captured by Cryo-Electron Microscopy. *Cell* **107**: 869-879.
143. Ranson NA, Clare DK, Farr GW, Houldershaw D, Horwich AL, & Saibil HR (2006) Allosteric Signaling of ATP Hydrolysis in GroEL-GroES Complexes. *Nature Struct & Mol Biol* **13**: 147-152.
144. Xu Z, Horwich AL, & Sigler PB (1997) The Crystal Structure of the Asymmetric GroEL-GroES-$(ADP)_7$ Chaperonin Complex. *Nature* **388**: 741-750.
145. Zheng W, Brooks BR, & Thirumalai D (2007) Allosteric Transitions in the Chaperonin GroEL Are Captured by a Dominant Normal Mode That Is Most Robust to Sequence Variations. *Biophys J* **93**: 2289-2299.
146. Zheng W, Brooks BR, & Thirumalai D (2006) Low-Frequency Normal Modes That Describe Allosteric Transitions in Biological Nanomachines Are Robust to Sequence Variations. *Proc Natl Acad Sci U S A* **103**: 7664--7669.
147. Kass I & Horovitz A (2002) Mapping Pathways of Allosteric Communication in GroEL by Analysis of Correlated Mutations. *Proteins* **48**: 611-617.
148. Chen Y, Reilly K, & Chang Y (2006) Evolutionarily Conserved Allosteric Network in the Cys Loop Family of Ligand-Gated Ion Channels Revealed by Statistical Covariance Analyses. *J Biol Chem* **281**: 18184-18192.
149. de Groot BL, Vriend G, & Berendsen HJC (1999) Conformational Changes in the Chaperonin GroEL: New Insights into the Allosteric Mechanism. *J Mol Biol* **286**: 1241-1249.
150. Ma J & Karplus M (1998) The Allosteric Mechanism of the Chaperonin GroEL: A Dynamic Analysis. *Proc Natl Acad Sci U S A* **95**: 8502-8507.
151. Sewell BT, Best RB, Chen S, Roseman AM, Farr GW, Horwich AL, & Saibil HR (2004) A Mutant Chaperonin with Rearranged Inter-Ring Electrostatic Contacts and Temperature-Sensitive Dissociation. *Nature Struct & Mol Biol* **11**: 1128-1133.

152. Shewmaker F, Kerner MJ, Hayer-Hartl M, Klein G, Georgopoulos C, & Landry SJ (2004) A Mobile Loop Order-Disorder Transition Modulates the Speed of Chaperonin Cycling. *Protein Sci* **13**: 2139-2148.
153. Stan G, Thirumalai D, Lorimer GH, & Brooks BR (2003) Annealing Function of GroEL: Structural and Bioinformatic Analysis. *Biophys Chem* **100**: 453-467.
154. Hyeon C, Lorimer GH, & Thirumalai D (2006) Dynamics of Allosteric Transitions in GroEL. *Proc Natl Acad Sci U S A* **103**: 18939-18944.
155. Horovitz A, Amir A, Danziger O, & Kafri G (2002) Φ Value Analysis of Heterogeneity in Pathways of Allosteric Transitions: Evidence for Parallel Pathways of ATP-Induced Conformational Changes in a GroEL Ring. *Proc Natl Acad Sci U S A* **99**: 14095-14097.
156. Clare DK, Bakkes PJ, van Heerikhuizen H, van der Vies SM, & Saibil HR (2006) An Expanded Protein Folding Cage in the GroEL-Gp31 Complex. *J Mol Biol* **358**: 905-911.
157. Kabsch W (1978) A Discussion of the Solution for the Best Rotation to Related Two Sets of Vectors. *Acta Crystallogr.* **34**: 827-828.
158. Eyal E, Yang L-W, & Bahar I (2006) Anisotropic Network Model: Systematic Evaluation and a New Web Interface. *Bioinformatics* **22**: 2619-2627.
159. Mouawad L & Perahia D (1996) Motions in Hemoglobin Studied by Normal Mode Analysis and Energy Minimization: Evidence for the Existence of Tertiary T-Like, Quaternary R-Like Intermediate Structures. *J Mol Biol* **258**: 393-410.
160. Petrone P & Pande VS (2006) Can Conformational Change Be Described by Only a Few Normal Modes? *Biophys J* **90**: 1583-1593.
161. Yang L, Song G, & Jernigan RL (2007) How Well Can We Understand Large-Scale Protein Motions Using Normal Modes of Elastic Network Models? *Biophys J* **93**: 920-929.
162. Rueda M, Chacón P, & Orozco M (2007) Thorough Validation of Protein Normal Mode Analysis: A Comparative Study with Essential Dynamics. *Structure* **15**: 565-575.
163. Teotico DG, Frazier ML, Ding F, Dokholyan NV, Temple BRS, & Redinbo MR (2008) Active Nuclear Receptors Exhibit Highly Correlated Af-2 Domain Motions. *PLoS Computational Biology* **4**: e1000111.
164. Yifrach O & Horovitz A (1994) Two Lines of Allosteric Communication in the Oligomeric Chaperonin GroEL Are Revealed by the Single Mutation Arg196-->Ala. *J Mol Biol* **243**: 397-401.
165. Yifrach O & Horovitz A (1998) Mapping the Transition State of the Allosteric Pathway of GroEL by Protein Engineering. *J Am Chem Soc* **120**: 13262-13263.
166. Horovitz A & Yifrach O (2000) On the Relationship between the Hill Coefficients for Steady-State and Transient Kinetic Data: A Criterion for Concerted Transitions in Allosteric Proteins. *Bull Math Biol* **62**: 241-246.
167. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, & Ben-Tal N (2003) ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* **19**: 163-164.
168. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, & Ben-Tal N (2005) ConSurf 2005: The Projection of Evolutionary Conservation Scores of Residues on Protein Structures. *Nucl. Acids Res.* **33**: W299-302.
169. Fenton WA, Kashi Y, Furtak K, & Norwich AL (1994) Residues in Chaperonin GroEL Required for Polypeptide Binding and Release. *Nature* **371**: 614-619.

170. Saibil HR (1996) Chaperonin Structure and Conformational Changes in *The Chaperonins*, ed. Ellis RL (Academic Press, London, UK), p. 261.
171. van der Vaart A, Ma J, & Karplus M (2004) The Unfolding Action of GroEL on a Protein Substrate. *Biophys J* **87**: 562-573.
172. Sun Z, Scott DJ, & Lund PA (2003) Isolation and Characterisation of Mutants of GroEL That Are Fully Functional as Single Rings. *J Mol Biol* **332**: 715-728.
173. Lange OF, Lakomek N-A, Fares C, Schroder GF, Walter KFA, Becker S, Meiler J, Grubmuller H, Griesinger C, & de Groot BL (2008) Recognition Dynamics up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* **320**: 1471-1475.
174. Ivanovska IL, de Pablo PJ, Ibarra B, Sgalari G, MacKintosh FC, Carrascosa JL, Schmidt CF, & Wuite GJL (2004) Bacteriophage Capsids: Tough Nanoshells with Complex Elastic Properties. *Proc Natl Acad Sci U S A* **101**: 7600-7605.
175. Howard J (2001) *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Associates, MA).
176. Tachibana M, Kojima K, Ikuyama R, Kobayashi Y, & Ataka M (2000) Sound Velocity and Dynamic Elastic Constants of Lysozyme Single Crystals. *Chem Phys Lett* **332**: 259-264.
177. Koizumi H, Tachibana M, & Kojima K (2006) Observation of All the Components of Elastic Constants Using Tetragonal Hen Egg-White Lysozyme Crystals Dehydrated at 42% Relative Humidity. *Phys. Rev. E* **73**: 041910
178. de Pablo PJ, Schaap IA, MacKintosh FC, & Schmidt CF (2003) Deformation and Collapse of Microtubules on the Nanometer Scale. *Phys Rev Lett* **91**: 098101.
179. Stephanidis B, Adichtchev S, Gouet P, McPherson A, & Mermet A (2007) Elastic Properties of Viruses **93**: 1354-1359.
180. Kundu S, Melton JS, Sorensen DC, & Phillips GN, Jr. (2002) Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models. *Biophys J* **83**: 723-732.
181. Bernard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, & Martin Karplus (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **4**: 187-217.
182. Case DA, Darden TA, T.E. Cheatham I, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W*, et al.* (2008) AMBER 10:
183. Chandler D (1978) Statistical Mechanics of Isomerization Dynamics in Liquids and the Transition State Approximation. *J. Chem. Phys.* **68**: 2959-2970.
184. Best RB & Hummer G (2005) Reaction Coordinates and Rates from Transition Paths. *Proc Natl Acad Sci U S A* **102**: 6732-6737.